TECHNISCHE UNIVERSITÄT DARMSTADT

Novel Interaction Concepts for Event Participation Through Social Television

Dissertation
Zur Erlangung des akademischen Grades
eines Doktor rerum naturalium (Dr. rer. nat.)

Eingereicht von
MSc. Niloofar Dezfuli
geb. in Teheran

Angenommen vom
Fachbereich Informatik der Technischen Universität Darmstadt

Gutachter: Prof. Dr. Max Mühlhäuser (TU Darmstadt, Germany)
Prof. Dr.Dick Bulterman (FX Palo Alto Laboratory, Inc., USA
Vrije Univ. Amsterdam, Netherlands)

Tag der Einreichung: 06.11.2014
Tag der Disputation: 18.12.2014

D17–Darmstadt 2015

Abstract

The present thesis is concerned with Human-Computer Interaction in the context of what we call "social television": the recent embracement of television systems, internet technologies, and social services. The ultimate research goal of this thesis is the development of novel, appropriate *interaction concepts* to support social television users – located either in situ (mobile in the field) or at-home – dealing with both user-generated or authoritative (professional) contents.

To develop concrete interaction concepts and to empirically validate their suitability, the present thesis puts a narrow focus on the event participation where spectators share the very same location and happening. Event participation is certainly a great way for many to get involved in a fun activity and to form a community. The so-called local-scope mass events – such as stadium -based sporting events – attract a large number of spectators to witness and experience the live atmosphere of something extraordinary in-situ. Moreover, the availability of a variety of event-related multimedia content – such as professional broadcasts and user-generated content – enables an even larger number of TV viewers to follow live happenings remotely at homes.

Beside all inherent advantages that each type of the event participation offers, they suffer from several limitations. The monotony of a single (and thus restricted) viewing perspective, particularly in all-seater event venues, may potentially result in missing important moments, scenes and sub-events that occur out of the view of or far away from spectators. As another consequence of being physically restricted to a particular seat, social interaction of spectators is limited to nearby people. On the other hand, TV viewers following the event remotely require more immediate and less distracting interaction modalities at homes. Moreover, they may feel socially disconnected as there is less proper and immediate way to share their excitement, opinion, and support for their team which may mar event experiences. The present thesis aims at reducing these deficiencies by leveraging late technology trends, in particular mobile video sharing and body-centric sensing approaches. The contributions of the present thesis are placed alongside three main research directions.

The first, in-situ experiences, explores how mobile user-generated content (particularly video) sharing in real-time can support the coconstruction of experiences during local-scope mass events. It particularly focuses on investigating the design requirements and guidelines for mobile systems supporting both multicamera viewing perspective and social interactions in-situ. Based on an iterative design process, it further contributes a set of novel interaction techniques for user-generated live

video sharing which are then evaluated in two field studies. The results indicate that the novel digital experience with live video on mobile devices enhances the overall event experiences in-situ.

The second research direction, at-home experiences, looks at how viewers can interact with televisions in a less-distracting way, particularly while watching a live program. It focuses on device-less and body-centric interaction that goes beyond the traditional button-based remote control paradigm. In this light, two novel body-based TV user interfaces are proposed supporting the whole body and hand-based inputs. The findings of a set of user studies confirmed that leveraging the human body as an interface for the television has various advantages – such as being omnipresent, device-less, and eyes-free – and thus can enhance the experience of TV watching activity in living room settings.

Finally, the third research direction, home-field (connected) experiences, addresses how the gap between people participating the event in the field and following it from home can be bridged. The social patterns and preferences of TV viewers for watching – not only live coverage of mass events but also other main TV genres – with non-collocated people are initially examined. The analysis leads to a set of requirements that served as rationale for the design of interaction techniques to connect sport fans in both realms. This is achieved through sharing mobile live video sharing as well as implicit gestural information of viewers in front of the TV. An early user feedback evaluation reveals a great potential for mutually contributing to the event engagement, potentially leading to more immersive and socially connected experiences during live sporting events.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit Mensch-Computer Interaktion im Kontext von, wie wir es nennen, sozialem Fernsehen: Die Vereinigung von Fernsehgeräten, Internettechnologien und sozialen Onlinediensten. Das ultimative Forschungsziel dieser Arbeit ist die Entwicklung von neuartigen, angemessenen Interaktionskonzepten für Nutzer von sozialem Fernsehen – sowohl in-situ (mobil im Feld) oder zu Hause –, die sowohl nutzergenerierte als auch offizielle (professionelle) Inhalte parallel nutzen.

Um konkrete Interaktionskonzepte zu entwickeln und deren Eignung empirisch zu validieren, legt diese Arbeit einen engen Fokus auf die Teilnahme an Ereignissen, bei denen Zuschauer sowohl Ort als auch Veranstaltung exakt gemeinsam haben. Die Teilnahme an Ereignissen ist für viele ein guter Weg, Spaß zu haben und Gemeinschaft zu formen/erleben. Die sogenannten lokalen Großereignisse – wie Sportveranstaltungen im Stadion – ziehen eine große Anzahl Zuschauer an, um die Atmosphäre etwas außergewöhnlichen in-situ zu erleben. Zusätzlich erlaubt es die Verfügbarkeit einer Vielzahl an ereignisbezogenen Multimediainhalten – wie professionelle Übertragungen und nutzergenerierte Inhalte – einer noch größeren Menge an Zuschauern dem Ereignis aus der Ferne – z.B. zu Hause – beizuwohnen.

Neben allen Vorteilen die verschiedenen Formen der Ereignisteilnahme bieten, haben diese auch einige Einschränkungen. Die feste (und damit eingeschränkte) Sichtperspektive, insbesondere bei Ereignissen mit Sitzplätzen, kann insbesondere dazu führen, dass man Details wichtiger Augenblicke, Szenen und Unterereignisse, verpasst, die außerhalb des Sichtbereichs oder weit weg vom Zuschauer passieren. Eine weitere Folge des festen Sitzplatzes ist, dass die soziale Interaktion zwischen den Zuschauern beschränkt ist auf Zuschauer in der Nähe. Auf der anderen Seite benötigen Fernsehzuschauer, die das Ereignis aus der Ferne von zu Hause verfolgen, direktere und weniger ablenkende Interaktionsmodalitäten. Zusätzlich können sie sich sozial isoliert fühlen, da es weniger direkte Wege gibt, ihre Aufregung, Meinung und Unterstützung für ein Team mit anderen zu teilen, was das Erlebnis verderben kann. Ziel dieser Arbeit ist es, diese Nachteile durch Nutzung aktueller Technologietrends, insbesondere mobilen Videostreamings und körperzentrierter Eingabe zu reduzieren. Die Beiträge dieser Arbeit gliedern sich in drei Hauptforschungsrichtungen:

Die Erste, in-situ Erlebnisse, untersucht wie Teilen von mobil benutzergeneriertem Inhalt (insbesondere Video) die „Cokonstruktion" von Erlebnissen während lokalen Großereignissen unterstützen kann. Sie untersucht dabei insbesondere die Designanforderungen und Richtlinien für mobile Systeme, die in-situ sowohl mehrere

Kameraperspektiven als auch soziale Interaktion unterstützen. Basierend auf einem iterativen Designprozess, trägt sie weiterhin eine Menge neuartiger Interaktionstechniken für das Teilen benutzergenerierter Livevideos bei, welche in zwei Feldstudien evaluiert wurden. Die Ergebnisse zeigen dass das neue digitale Erlebnis mit Livevideo auf Mobilgeräten das in-situ Gesamterlebnis verbessern.

Die zweite Forschungsrichtung, „zu Hause“ Erlebnisse, untersucht, wie Zuschauer, insbesondere während Liveübertragungen, weniger ablenkend mit ihrem Fernseher interagieren können. Sie fokussiert sich auf gerätelose und körperbasierte Interaktion, die über das traditionelle tastenbasierte Fernbedienungsprinzip hinausgeht. In diesem Licht werden zwei neue körperbasierter Fernseher-Benutzerschnittstellen vorgeschlagen, die ganzkörper- und handbasierte Eingabe unterstützen. Die Ergebnisse einer Reihe von Nutzerstudien bestätigten, dass die Nutzung des Körpers als Schnittstelle für die Fernseherbedienung verschiedene Vorteile bietet, unter anderem ist sie allgegenwärtig, gerätelos und eyes-free und kann daher das Fernseherlebnis im Wohnzimmer verbessern.

Die dritte Forschungsrichtung, daheim– in-situ (verbundene) Erlebnisse – beschäftigt sich damit wie die Lücke zwischen Menschen, die vor Ort sind und denen die von zu Hause zusehen, geschlossen werden kann. Die sozialen Muster und Vorlieben von Fernsehzuschauern die gemeinsam – nicht nur Liveübertragungen von Großereignissen, sondern auch andere Fernsehgenres – an verschiedenen Orten sehen werden zunächst untersucht. Die Analyse führte zu einer Reihe von Anforderungen, die als Grundlage für das Design der Interaktionstechniken zur Verbindung von Sportfans in beiden Bereichen durch bidirektionales mobiles Livevideostreaming und implizite Gesteninformation der Fernsehzuschauer diente. Erstes Nutzerfeedback zeigte ein großes Potential für einen beidseitigen Beitrag zum Ereignis, welcher potentiell zu einem immersiveren und sozial verbundeneren Erlebnis während Sportereignissen führt.

Acknowledgments

Even though the words and thoughts are my own, the Ph.D. dissertation you are about to read is not what I could create without the support and help of many people. First, I want to thank people who crossed my path during this creation process and have helped shape this research. In particular, I thank my adviser and mentor Prof. Max Mühlhäuser (Technische Universität Darmstadt, Germany). It has been an honor to be his student and have his great support and inspiring discussions he provided throughout the years of my research.

I am also very grateful to Prof. Dick Bulterman (President of FX Palo Alto Laboratory, USA) for acting as co-referee and I appreciate all his contributions of time and ideas to make my Ph.D. productive and stimulating.

I was very fortunate to work in a friendly, fruitful, and fun environment, i.e. Telecooperation (TK) Lab at Technische Universität Darmstadt, in which I collaborate with an excellent group of colleagues: Simon Olberding, Thorsten Strufe, Daniel Schreiber, Felix Heinrichs, Elke Halla, and Nina Jäger. I want to thank all other former and current TK members that made my time there such a wonderful experience. Likewise, I sincerely thank all the students I worked with for their valuable creative contributions: Murat Özkorkmaz, Johan Raspasovic, Manolis Pavlakis, and Sebastian Günther.

I am most grateful to my awesome fellow research group members and–most importantly–friends: Jochen Huber, Roman Lissermann, Florian Müller, Jan Riemann, and Martin Schmitz.

Words cannot express how especially grateful I am to my mother, Minoo, and father, Saeid, for all of the sacrifices that you have made on my behalf. Your prayer for me was what sustained me thus far. I would also like to thank my brother, Siavash, and his lovely wife, who always supported me, and incented me to strive towards my goal.

At the end, I would like express appreciation to my beloved husband, Mo, and my little angel, Lena, who spent sleepless nights with and were always my support in the moments when i needed it most and there was no one to answer my queries. I will be grateful forever for your love.

Contents

Chapter 1

Introduction

Contents

The thesis present hereafter is concerned with human-computer-interaction (HCI) in the context of what we call "social television" (STV, see below). Three facets of STV are investigated:

- social television in situ for user-generated content
- social television at home for authoritative and professional content ("classical broadcast")
- the bridging of in situ and at-home social television for mixed content

For all these facets, the research goal of the present thesis is the development of novel, appropriate *interaction concepts*. This research is empirically grounded in both ex-ante exploratory studies and ex-post evaluation of the interaction concepts developed.

In order to develop concrete interaction concepts and to empirically validate their suitability, the present thesis puts a narrow focus on what we call local-scope mass events, ranging from festivals and carnivals to stadium sports events (see details below).

The first section below discusses the motivation for the present thesis. Thereby, we first share a few general considerations before turning to the thesis focus (i.e., local-scope mass events). We highlight challenges and major deficiencies of in situ and at-home participation (as a spectator) from the viewing- and social-experience perspectives.

The second section summarizes the major scientific contributions of the thesis, structured according to the above, which mentioned three facets of STV, equal to the three research directions of this thesis. Section 1.3 provides an overview of the research methodology applied throughout the thesis. Finally, section 1.4 guides the reader through the remainder of the present thesis.

1.1 Motivation

The recent embracement of Internet technologies and television (TV) systems has drastically changed the way TV is incorporated into the daily routine of its audiences. This new television, which we refer to as social TV (STV) throughout this thesis, is a result of maturing and converging technologies in various fields, such as multimedia streaming, wireless or wired telephony, and telecommunication, combined with broadband connectivity. It promises to deliver a world of content, services, and applications to any device–ranging from stationary TVs or personal computers to laptops, tablet, and smartphones–used either at homes or on the go at anytime [O'Hara 2007].

In terms of services and contents, STV offers much more than watching a number of limited professional broadcasts. The ever-growing number of subscribers to online social networks, such as Facebook and MySpace, opens up a tremendous opportunity to extend social interactions centered around the TV to a global scale (compared to the local collocated social interactions of traditional TV systems). Moreover, STV allows viewers to access social TV applications (apps) via integrated app store markets, providing a range of online social media services right on the TV screen. One important consequence of this is the possibility of watching user-generated content from video-sharing platforms (such as YouTube or Vimeo) besides to the professionally crafted contents. The immense growth in user-generated contents (e.g., more than 100 hours of video are posted in one minute in YouTube [1]) can potentially enrich the viewing experience.

With the rapid growth of the number of broadband links to home and the available bandwidth to mobile devices, the connectivity of TV viewers is constantly increasing [O'Hara 2007]. This provides a basis for the second promise of STV: going beyond the home delivery model of the traditional television (and also most current deployment of STV systems) and extending content delivery to potentially everywhere. Thus, STV experiences are not bounded to living-room environments anymore. People in the field or those on mobile can have access to the world of other

[1] http://www.youtube.com/t/press_statistics/

TV viewers, contents, and the activities of social media as additional information right on their handheld mobile computing devices (smartphones and tablets).

Through such a social engagement and interactivity with TV content on multiple devices, the TV experience, which is traditionally characterized as a passive and lean-back experience, has become active. As a consequence, STV adds interactivity to the TV, in which the content itself or the presentation manner of the content or even the presentation order of the content can be affected by the viewer [Jääskeläinen 2001]. Regardless of their location, viewers can exchange comments while watching the same content, suggest other shows and programs, collaboratively vote for a quiz answer, and even setup an on-screen video or voice call.

While STV and related technologies and services enable a richer content to be delivered to TV viewers located anywhere at any time, at the same time they impose fundamental challenges on both the technical and deployment levels as well as on the interaction design and user experience level. While technical challenges have been addressed to a great extent–as several STV systems have been widely deployed in a number of countries [Edwardson 2014]–interaction and user-experience-related challenges are still under exploration. This is mainly due to the drastic changes in the traditional TV viewing experience, which are becoming active (compared to passive), social, and networked (compared to the collocated local social origin of TVs), happening in diverse settings, such as in public spaces, on the move, and at home (compared to only at home usage).

Thus, the overarching goal of this thesis is the development of novel appropriate interaction concepts, supporting social connectivity, and providing truly natural and immersive shared experiences for STV. These concepts are concretely designed and evaluated in the context of local-scope mass events (e.g., sporting events), supporting the event participation both in situ and at home. Given the nature of live events that are prime social- and media-based phenomena, the ultimate goal of this thesis is to improve not only the efficiency of interaction but also social connectivity to yield delightful and immersive user experiences while consuming media created both professionally or amateurish (user-generated).

1.1.1 Event Experiences

Local-scope mass events are prime social, economic, and media-intensive phenomena, implying large groups of spectators and visitors gathered in an immense spatial area to coexperience something extraordinary [Jacucci 2005]. Recently, live events have become a huge business, attracting billions of fans around the world. Among others, sport events, festivals (e.g., music, food, etc.), and carnivals are chief instances of large-scale events. These happenings attract not only a large number of spectators to directly witness and experience the life atmosphere but also an even larger number of remote viewers who follow the very specific topic through media coverage, like professional broadcasts or user-generated content. For example, the recent World Cup 2014 held in Brazil sets both visiting as well as viewing records [2]: 3.4 million visitors attended the matches in stadiums, and over 17 million viewers watched matches only in the United States. These numbers truly show the ever-growing popularity of following live events.

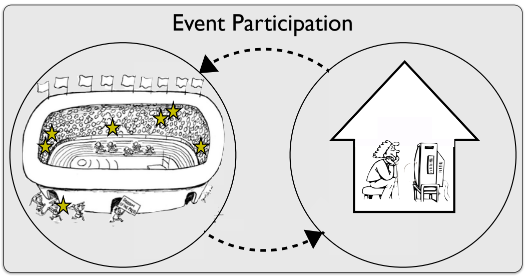


Figure 1.1: Research focus

As stated above, such events are generally experienced in two distinct ways based on the location of fans: in situ (at stadiums, arenas, and amphitheaters) and remote (where followers view professionally crafted content of the event) (cf. figure 1.1). The latter considers watching the event at living rooms in front of TV sets, at public places in front of large screens, or on mobile devices while on the go. Due to their popularity and frequent occurrences, this thesis focuses on addressing challenges for two primary ways of experiencing an event, *in situ* and *living room*, which we discuss below.

In situ Challenges

People attend mass public events to directly witness a lively atmosphere of something extraordinary. In addition, external interaction and socialization, novelty and uniqueness, as well as entertainment and excitement are listed as the main motivations of event-goers in literature [Trail 2005, Wann 1996]. Consequently, these

[2] http://www.fifa.com/worldcup/statistics/

events usually host a large number of spectators (over hundreds of thousand) who are spatially distributed within an immense stadium- or arena-size area. They can be centered around a single theme (e.g., sport events) or consist of different subevents taking place in parallel or one after the other.

Beside all inherent advantages that the live participation of an event offers, it also poses a number of challenges for spectators:

1) The first major problem associated with local-scope mass events is that spectators cannot easily perceive the whole event [Jacucci 2007b]. This is mainly due to the amplitude nature of events that impedes obtaining a real-time overview and awareness about the event, its (sub)events, and activities of fellow spectators. This problem becomes even more severe in events' environments where spectators are assigned to a particular seat and thus are limited to a certain viewing angle.

2) The second challenge is the limited social interaction formed in situ. One of the main characteristic of large-scale events is to socially construct an extraordinary experience. As matter of fact, a large number of spectators visit the event in groups constituting of a number of people who know each other in one way or another (e.g., friends, families members, colleagues, etc.) [Jacucci 2007b]. This can potentially limit the boundaries of social interaction and collaboration to individual groups of already known friends. Prior studies have shown that spectators have seldom directed attention to social interaction with other fellow spectators (particularly strangers) and involved in other's activities [Sun 2007].

With the mass adoption of Internet-enabled smartphones, nowadays, almost all spectators have a mobile access to the Internet and thus to a myriad of online information and services. Since most of smartphones are equipped with an integrated high-resolution camera, spectators frequently capture exciting moments and subhappenings in the form of photos or videos. Such audiovisual user-generated content is usually shared via spectators' online social networks or directly sent to their friends to foster conversation around the topic. This practice (i.e., *mobile media sharing*) has recently become a well-established and common practice in local-scope mass events [Jacucci 2007a]. It is, however unclear how such a rich source of on-site information can be used to address aforementioned challenges in the section 1.1.1. It can potentially provide efficient overview, increase real-time awareness, and act as a means to initiate social interaction among spectators through the deign of immediate appropriate interactions, thus contributing to the active participation. Therefore, one stream of research in this thesis investigates:

- How mobile user-generated content (particularly video) sharing in real time can support the coconstruction of experiences in situ during local-scope mass events?
- How mobile live video sharing can be harnessed for enhancing the event experience?
- What are the design requirements for mobile systems supporting both viewing and social experiences in situ?

Living Room Challenges

As we mentioned above, there exist a larger set of people that experience a live event from their homes remotely. On one hand, compared to the in situ atmosphere that is highly dynamic and live, following the event at homes is typically indicated as a *lean-back* practice. On the other hand, in addition to the live professional broadcasts, viewers can access various sources of online information related to the event as well as watch user-generated videos or photos live posted from the event, all on the TV screen. Furthermore, STV services and applications open up opportunities to communicate with other remotely located viewers as well as spectators in the field and vice versa. While these features support viewers to get the most out of the event from a distance, they give a rise to myriad of challenges that are needed to be addressed for seamless and successful TV experiences:

1) One salient challenge is the user interaction with STVs. Typically, TV viewer input on television is supported through remote controls. Common examples are conventional remotes with physical buttons or touch-based interfaces on smart phones. In effect, viewers are always required to use a particular mediator device to interact with TVs. While this is a well-established interaction paradigm, it has various drawbacks. The device itself can be easily out of reach or misplaced. Moreover, touch-based interfaces require viewers to frequently switch their attention between the TV and smartphones screens that may eventually distract the viewers.

Moreover, such living room experiences traditionally constitute of one or several event fans (viewers) gathering together in front of a television set to watch professionally produced broadcasts. In this social context, it is always challenging to determine who owns the remote control and is responsible for the selection of television content at home [Lull 1982]. Therefore, the family interaction and content selection processes can influence watching experiences. These problems are more acute in case of following a live event when viewers are not willing to miss even a single moment of it. Thus, novel interaction supports are needed for STVs that are immediate (efficient), intuitive (highly easy to use and learn), and less distracting.

As the second stream of research, this thesis explore novel interaction concepts that go beyond the device-based remote control paradigm. More precisely, we investigate:

- How can viewers interact with TV in a deviceless and less distracting way that goes beyond the remote control paradigm?
- How can a deviceless interaction mode be designed and incorporated into the TV user-interface design?

2) The second at-home challenge originates from the integration of social networks and communication services with TV systems. Using such features, viewers can communicate with mobile spectators in the field in various forms (e.g., text messaging, video-audio communication, etc.), thus potentially contributing to the event remotely. At first glance, communication between homes and the field during live events might seem palatable and desirable to viewers, but inadequate design of interfaces and communication forms can significantly diminish the user experience and turn them into distracting and impractical than helpful.

Therefore, the third stream of research presented in this thesis addresses the challenge of appropriately bridging the gap between living rooms and the event venue by designing practical and effective communication channels. More precisely, we aim to systematically answer the following:

- How effectively can the gap between people in the field and at home be bridged (particularly through sharing real-time user-generated media)?

In sum, this section presented emerging challenges for both spectators and viewers who are interested in real-time following of large-scale events. It goes without saying that addressing these challenges will result in more immersive and delightful event experiences. Thus, the present thesis contributes novel interaction and user-experience concepts to support in situ, remote, and connected experiences for large-scale events.

1.2 Research Directions and Contributions

In response to each of these challenges of live events in both settings, three research directions (RD) are investigated; each includes a number of contributions that are presented in detail below.

RD1. Exploring Spectatorship Experiences In Situ

This research direction considers enhancing in situ user experiences for spectators. Particularly, we address decreased viewing and social experiences imposed by the physical restriction of local-scope mass events. We contribute a set of novel location-aware mobile video sharing concepts, which we call *CoStream*.

CoStream supports the coconstruction of experiences in situ. Its design is empirically grounded by following an iterative design process in which we conduct three focus group sessions. Based on findings, we contribute a set of interaction concepts and techniques for live user-generated video sharing and prototypically implement them in a coherent system. CoStream concepts are designed to provide in situ awareness as well as to foster the active participation and social interactions among spectators.

Further in this direction, we contribute results of two field studies in which the CoStream concepts and system are systematically evaluated. The studies are conducted during two soccer matches with a particular emphasize on the impact of CoStream on both social and event experiences. Results indicate that CoStream effectively supports the co construction of shared experiences and complements the overall in situ experiences of large-scale events.

RD2. Supporting Interaction for Living Room Experiences

In this direction, we target living room experiences and the challenges that viewers are faced with while interacting with television systems. We start our investigations by gaining a better understanding about people's use of body in front of the TV. Therefore, we conduct a field study to particularly look at how people spatially situate themselves in front of the TV and how they engage in watching activity. Findings of the study provide empirical foundations for the design of two novel body-based TV user interfaces concepts, namely *CouchTV* and *PalmRC*, as the main contributions of this research direction.

CouchTV concept supports various course-grained interactions with TV systems that rely entirely on the spatial and postural information of viewers. It contributes novel interaction techniques for (re)engaging in TV watching activity, providing

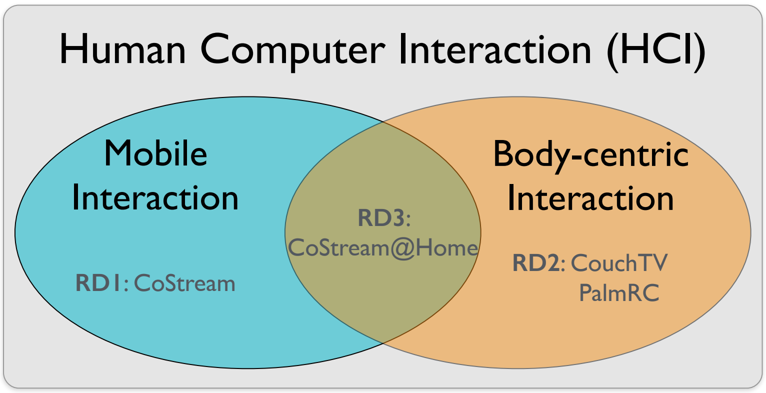


Figure 1.2: The context of the contributions made in this thesis.

appropriate level of awareness, and displaying supplementary information related to live event program. The CouchTV interface is evaluated in an initial user feedback session with 12 groups of TV viewers.

In PalmRC, we appropriate the palm of the hand as a means to enable fine-grained interactions with the TV. It is a novel eyes-free input style for television systems that allows TV viewers to perform spatial interactions with empty hands through the sense of proprioception. The PalmRC interface concepts are evaluated through a series of user studies focusing on the effectiveness and user experience of this novel TV input modality.

RD.3 Connecting Event Experiences

The third research direction investigates how the experiential distance between TV viewers at homes and spectators in the field can be effectively bridged. To gain a better understanding of noncollocated event participation, we first present results of a field study in which we examine social interpersonal relationship patterns and preferences of TV viewers for remote watching while following not only live coverage of local-scope mass events such as sport matches but also all other main TV genres. Based on the study findings, we contribute a set of concepts for connecting in situ and living room experiences through bidirectional mobile video sharing, which we call CoStream@Home.

Based on the CoStream@Home concept, we exploit mobile devices in both realms as a means for mutually contributing to the event engagement and thus providing more immersive and socially connected experiences during large-scale events. In addition to the video sharing communication, we stimulate social interactions by proposing a real-time communication channel, based on gestural and emotional information (cheering, frustration, etc.) of viewers in front of the TV. We further illustrate how we envision this to aid in bridging the aforementioned gap and believe that such information can open up novel social interaction appropriate for both realms. In an early user feedback session, we test the CoStream@Home concept and its interaction techniques with five experts in the field of HCI.

1.3 Research Context and Methodology

In terms of the general scientific disciplines, the contributions of this thesis are primary situated in the filed of HCI (i.e., the study of interaction between people and computers and the design of novel interface approaches). Within this field, the first research direction explicitly contributes to the field of mobile interaction (i.e., studies and systems specifically focusing on user interface concepts for mobile devices to be used on the go (see [Huber 2012] for an extensive review). The second research direction contributes to the field of body-centric user interfaces, where the human body is used as an interactive platform for both input and output (see [Wagner 2013] for an extensive review). The context of the third research direction is situated at the intersection of the both fields. Figure 1.2 illustrates the research context of this thesis.

The contributions of this thesis detailed above are made through sound and well-known research methodologies. They are empirically grounded by exploring the natural behavior of users and examining literature in the respective field. Then based on empirical qualitative and quantitative findings, we follow a user-centered design approach [Norman 1986] in which we iteratively design and analyze interaction concepts with potential users. More precisely, each design iteration started with an initial study (in form of focus group or exploratory study) to systematically understand and analyze the current practices and behavior of users.

The findings of initial studies are compiled to a set of requirements to be considered in our concept and system designs. We then designed novel interaction concepts in a cyclic design process that were prophetically implemented. Finally, the concepts were evaluated with potential users in order to test the validity of the assumptions. In the evaluations (conducted both in the lab or in the field), we opted for both quantitative and qualitative data collection methods to measure both efficiency and user experience. A schematic view of the research method followed in this thesis is depicted in figure 1.3.

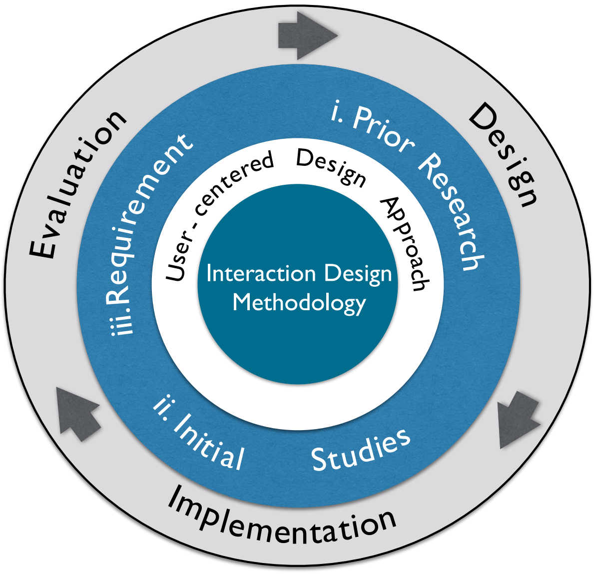


Figure 1.3: Overview over the research methodology.

1.4 Thesis Outline and Publications

This thesis is structured in three main chapters, where each includes one research direction. Chapter 2 focuses on exploring spectatorship experiences in situ and accordingly, present our contributions (CoStream) mentioned above. Based on several initial studies in the form of focus group, we define a set of design requirements in this chapter. These are then covered in CoStream: concepts and a system to support the coconstruction of experiences in situ. After detailing on the CoStream user interface, we further present results of its evaluations conducted in the field.

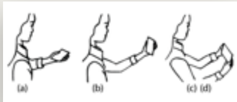
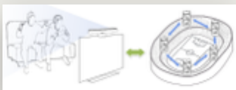
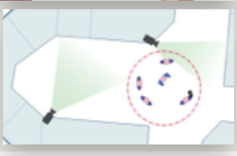
In Chapter 3, we approach viewership experiences that follow a live event remotely through television systems at homes. We start our investigation by observing how people situate themselves in front of the television and how this changes as the TV program goes on. We then turn our findings into a set of requirements that provide solid foundation for design of two main contributions of this chapter: CoachTV and PalmRC. Their design process, user-interface concepts and techniques, as well as evaluations are further documented in this chapter.

In Chapter 4, we address connecting live event experiences happening in the field and living rooms. More explicitly, we present CoStream@Home: a set of concepts and techniques to establish novel communicative ways between the two worlds. Its design is grounded by studying how people would like to virtually communicate and connect with other remote viewers and spectators, presented first in this chapter.

This is followed by describing CoStream@Home features and concepts as well as an evaluation in the form of an early user feedback session. Chapter 5 includes conclusions and possible future research directions.

Contents, ideas, and figures presented in the three main chapters have been published previously in the proceedings of international conferences such as, ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM Multimedia (ACM-MM), ACM Conference on Interactive Experiences for Television and Online Video (TVX - previously known as EuroiTV: European Conference on Interactive Television), as well as international workshops and scientific magazines.

Contents presented in Chapter 2 are published in CHI'12 [Dezfuli 2012a] and British HCI'13 [Dezuli 2013]. Contributions presented in Chapter 3 are published in the proceedings of EuroiTV'12 [Dezfuli 2012d, Dezfuli 2012c] and CHI'12 [Dezfuli 2012b] conferences as well as an article in Behaviour & Information Technology 2013 [Dezfuli 2014]. The study of interpersonal relationships is published in EuroiTV'12 [Dezfuli 2011]. Parts of the chapter on bridging the event experiences are published in SAM'13 (in conjunction with ACM MM'13) [Dezfuli 2013a] and CHI'13 [Dezfuli 2013b] workshops proceedings. Figure 1.4 provides a written and visual summary of which materials are used in the thesis chapters along with the corresponding publications.

Thesis Chapters	Corresponding Publications
1. Introduction	
2. Exploring Expectatorship Experiences In-Situ	 **Dezfuli**, , Huber, Olberding, Mühlhäuser, CoStream: In-situ Co-construction of Shared Experiences Through Mobile Video Sharing During Live Events, *In Proc. CHIEA 2012.*  **Dezfuli**, Huber, F. Churchill, Mühlhäuser. CoStream: Co-construction of Shared Experiences through Mobile Live Video Sharing, *In Proc BCS-HCI 2013.*
3. Supporting Interaction for Living Room Experiences	 **Dezfuli**, Pavlakis, Müller, Khalilbeigi and Mühlhäuser, CouchTV: Leveraging the Spatial Information of Viewers for Social Interactive Television Systems. *In Proc EuroITV 2012.* **Dezfuli** , Khalilbeigi , Huber , Müller , Mühlhäuser, PalmRC: imaginary palm-based remote control for eyes-free television interaction, *In Proc EuroITV 2012.* **(Best Paper Award)**  **Dezfuli**, Khalilbeigi, Huber, Müller, Mühlhäuser, Leveraging the Hand Surface as an Eyes-free TV Remote Control, *In Proc. CHIEA 2012.* **Dezfuli**, Khalilbeigi, Huber, Özkokmaz, Mühlhäuser. PalmRC: Leveraging the Palm Surface as an Imaginary Eyes-free TV Remote Control, *In Proc. Journal of Behaviour & Information Technology 2013.*
4. Connecting Shared Event Experiences	 **Dezfuli**, Khalilbeigi, Geerts, Mühlhäuser, A Study on Interpersonal Relationships for Social Interactive Television, *In Proc. EuroITV 2011.*  **Dezfuli** , Günther , Khalilbeigi , Mühlhäuser , Huber, CoStream@Home: connected live event experiences, *In Proc. SAM Workshop at ACM Multimedia 2013.* **Dezfuli**, Huber, Khalilbeigi and Mühlhäuser, Towards Connected Live Event Experiences, *In Proc. Exploring and Enhancing the UX for TV Workshop at CHI 2013.*
5. Conclusions	 Müller, Günther, **Dezfuli**, Sahami, Khalilbeigi and Mühlhäuser, Toward proximity-based hand input for on-body user interfaces. Submitted to CHI 2015.  Müller, Günther, **Dezfuli**, Khalilbeigi, Pavlakis and Mühlhäuser, CoCenter: developing an interactive 3D visualization for video monitoring centers. Submitted to ISCRAM 2015.

Chapter 2

Exploring Spectatorship Experiences In Situ

Contents

The explosive growth of capable mobile devices equipped with high-resolution cameras and superefficient mobile networks are dramatically changing the in situ experiences of spectatorship in local-scope mass events. Spectators are able to capture, share, and access user-generated contents virtually anywhere and anytime.

Over the past years, active video capturing has become an expectation and a norm, particularly during local-scope mass events such as sporting matches or musical concerts. In such events, spectators frequently capture videos snippets using mobile phones. Typically, these user-generated visual contents are further used in two main ways: either they are stored for personal archival as memories of the special moments of events or given the recent emergence of video sharing and social networking platforms (e.g., Facebook, YouTube, etc.), where they are uploaded and shared online with spectators' social circles. While these ways are certainly useful for extending the event's reach through time and making it searchable, we argue that *live* sharing of user-generated videos, as the third way to use such contents, can potentially promote coexperiencing the events in real time.

Recent research has shown that live and synchronous video sharing is a practical and useful way for bridging physical distances and supporting the "being there" [Hollan 1992]. Especially, this has shown to enhance experiences in events, like car rallies and carnivals, where spectators are scattered across different sub-events and thus, unable to witness the whole event. We believe that live and synchronous mobile video sharing among spectators of not only large-scale but also local-scope mass events, such as concerts, arena, and stadium sports, can provide fundamentally new experiences and more immersive social interactions for spectators.

This chapter focuses on investigating live user-generated video sharing for events' spectators that can obtain a high-level overview of the whole event through their peripheral vision but cannot perceive details of the event, easily. They may even be assigned to certain seats or not easily move to other location due to the crowd and thus are restricted to a particular viewing angle. As a consequent, the spectator's social interaction is limited either to accompanying friends or nearby spectators, instead of dispersing the experience to a broader audience.

In this chapter, to address these challenges (i.e., decreased (1) viewing and (2) social experiences imposed by physical restriction in such spaces) we propose a novel live mobile video sharing concept. More explicitly, leveraging smartphones as a communication means, we investigate if this has the potential to enrich the overall event experience where spectators share the same event and the same location. To do so, we contribute a set of interaction concepts and a mobile prototype system supporting live user-generated video sharing for in situ spaces that we call *CoStream*. The design of CoStream is empirically grounded in an iterative design process in which we conducted several focus group studies. We then evaluate the CoStream interaction concepts and prototype through two field studies conducted during soccer matches with a particular emphasize on its impact on both social and event experiences. We contribute their results indicating that CoStream effectively can support the coconstruction of shared experiences and helped complement the overall event experience.

In summary, the contributions of this chapter are the following:

- Iterative design process to empirically ground the design of mobile live video sharing concepts,
- A set of novel mobile interaction concepts for live video sharing experiences during local-scope mass events, coined as CoStream,
- Two field studies using CoStream system in stadium that explored the aforementioned research questions,
- Design guidelines for future interfaces supporting the live coconstruction of shared experiences during events.

The remainder of this chapter is structured as follows. In section 4.1, we discuss related work in the field of mobile TV and multimedia sharing focusing on enriching spectator experiences during an event. We then present our novel mobile live video sharing system, CoStream, and its iterative design process as well as the interaction concept in 2.2. We then report on the use of CoStream during two field studies that explored the aforementioned research questions in Sections 2.3 and 2.4. Finally, this chapter concludes with a discussion of the results of both studies in 2.5.

> **Contribution Statement:** Most of the work presented here is based on[Dezuli 2013, Dezfuli 2012a]. I am the first author of these publications. I have initiated and lead the project. My coauthors have also contributed significantly. My supervisors, Elizabeth F.Churchill, Jochen Huber and Max Mühlhäuser, have contributed to the design of the system and helped in writing the papers.

2.1 Related Work

To place our work in context, we review the state of the art in four main areas. First, we discuss prior studies that explored the use of user-generated messages, such as tweets, to better understand the topic and trends of an event. While not directly relevant, this vein of research support the design of the CoStream concept by underlying the importance of user-generated content and visualizing its geographic distribution to better report the events.

Second, we review prior studies that investigate how additional information, such as pictogram and text, can aid fostering awareness and a better overview of an event.

These studies provided additional awareness cues into media-sharing applications to enhance the experience of the spectatorship. We believe that reviewing this stream of research can provide insights on how to support the necessary awareness required by spectators while attending in situ live events .

Third, we discuss previous work that investigated active media engagement for the event. This field of research can extensively inspire the design of CoStream concept by providing insights into the active spectatorship practice for live events.

Fourth, we revisit prior research on mobile sharing of multimedia during events, particularly in real time, as the most relevant research topic to our contributions presented in this chapter.

In the following, we analyze related work in the aforementioned four areas. We look into the strengths and weaknesses of each concerning our main objectives and design goals in this chapter: support obtaining efficient overview as well as detail about the event, support real-time sharing of user-generated multimedia, support immediate interactions, and motivate active engagement during live events. A short summary of the main findings of literature analysis and systems is given at the end of this section. Note that in section 2.2.1, we will present more concrete goals and requirements that are covered in the design of CoStream system.

2.1.1 User-Generated Messages for Event Exploration

Considerable research investigated the practice of sharing user-generated text messages in the form of microblogs, such as tweets, around live events. This practice popularity is rising and people use it to discuss the events, which they are attending or watching through professional broadcasts on TVs. There are studies that focused on analyzing such user-generated messages and accordingly developing visual representations that can help support a high-level summary of an event and allow users to draw higher-level conclusions. While not directly related to our focus (i.e., live video sharing during in-situ events) we describe how sharing user-generated messages (microblogs) and its geographic distribution and visualization can support better understanding of the story of an event.

Vox Civita [Diakopoulos 2010] is a timeline-based visual analytic tool to enable extracting news value from public text messages posted on Facebook and Twitter while broadcasting large-scale events, such as televised debates and speeches. This tool extracts and displays the information that may be ignored due to overwhelming amount of contents and yet can aid to better report the events. The authors have evaluated their system with 18 journalists while the U.S. presidential State of the

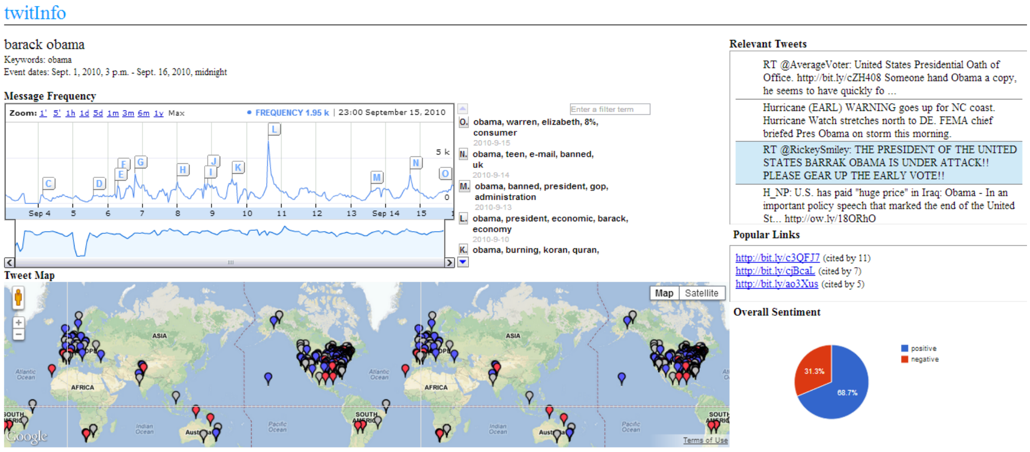


Figure 2.1: The main TwitInfo interface summarizing the event over time [Marcus 2011].

Union address debate of 2010. The study findings showed that journalists (and others) effectively use the tool to generate insight about the social media response to the event and about the event itself.

Similar in context, Shamma et al. [Shamma 2009] addressed the practice of sharing Twitter and its usage during the 2008 presidential debates. They demonstrated that the tweets can yield insights into the content and the story of the events. They showed that interesting events can be detected by looking at anomalies in the pulse of the sentiment signal in the event.

TwitInfo [Marcus 2011] presented an interface for visualizing and summarizing events on Twitter based on a timeline that allows users to navigate through a high-volume collection of tweets. The timeline enables users filtering the tweets to understand and describe important moments of events and highlight peaks of high tweet activity. The user interface also provides a focus+context visualization illustrating more peaks and sub-events. It also displays the geographic distribution of messages posted on Twitter as it is shown in figure 2.1. An evaluation of the system revealed that users were able to reconstruct meaningful summaries of events in a small amount of time.

Eddi [Bernstein 2010] is another examples of topic-browsing interfaces that enable tracking tweets streams related to a broadcast and televised event. Similar to the design of TwitInfo and Vox Civita, this interface also provided timeline-based visualizations of Twitter data and user's feed. However, it supported the temporal exploration of tweets only on a single topic that cannot be generalized to an arbitrary topic of interest.

Chakrabarti and Kunal [Chakrabarti 2011] presented a work to empirically tackle with noisy tweets to cover the sport events in real time. Their approach focused on recurring events, such as sport tournaments and leagues. They exploited the recurring characteristic of such events to enable their system to learn from previous matches in order to better summarize an ongoing event. They found that their approach can help to build models for the different subevents by using tweets from prior events, and consequently, learn the underlying hidden structure of such events.

In summary, the previous work in this stream of research explored understanding the flows and trends of events using shared user-generated short messages (e.g., tweets). We found that such contents are a valuable source of information that can be used for efficient and effective summarizing and understanding events. Moreover, visualizing tweets in the timeline manner and the geographical distribution enabled an appropriate representation of the event. Although these studies support the idea of leveraging shared user-generated content to understand and enrich the spectator and social experiences around events, these works are not focusing on real-time sharing and live communication, which may provide new access to a live event and generates novel experiences. Upon the support of the findings of the prior work discussed above, we investigate real-time sharing experiences during in situ events with a particular focus on mobile user-generated **video**.

2.1.2 Providing Additional Information During Events

In this section, we discuss the prior work that support sharing additional information to spectators attending the event or remote viewers watching it via TV broadcasts. These systems are specifically aimed at enhancing the different aspects of the spectatorship experience, particularly fostering awareness and obtaining a better overview over the event.

Holmquist et al. [Holmquist 1999] demonstrated a hand-held wearable device (called Hummingbird) that is designed to supply constant awareness information to users in any location. It provides visual and aural indications (i.e., visualize the identity of other users on a display and produce a sound) to members of a group when other members are close in large-scaled events. Hummingbird was evaluated in different event settings, such as a large rock music festival. The results reflected that Hummingbird can extend the awareness of presence and others' activities. The results of the study revealed that Hummingbird can also foster the feeling of connectedness among not only friends but also strangers in such large-scale and mass events where it can be hard to establish conversation with companions.

Figure 2.2: Mobile user interface of the TuVista system [Bentley 2009].

Bentley and Metcalf [Bentley 2008] designed three context-sharing applications as probes to explore how the transmission of contextual information can enhance the richness of communications. The applications were used to share experiences and feel connected to both people and an ongoing event in remote location. The first probe is a content-enabled phonebook that illustrates if any of the persons in the contact list is currently stationary or moving between places. The second probe is a music player that informs user if any of their closed contacts is playing a music. The last probe enables automatic video and photo sharing among close friends and families and receiving comments from them.

To investigate these probes, if they create fundamentally new social experiences, authors conducted three studies in which they examined each application. The authors found that people were quite willing to share their context in a wide variety of circumstances and for many purposes such as creating social awareness, helping others, continuing conversation, etc. Their findings also revealed that since this technology (i.e., mobile presence services) allows continual connection, it provides and increases constant awareness and enable people to feel in touch with each other even when no explicit message is exchanged.

eStadium [Ault 2008] provided awareness over an ongoing sport event by delivering a list of related on-demand videos retrieved from the Internet to mobile devices that enabled replay viewing. This mobile web application also provided real-time statistics and venue information to spectators to enhance the live event experience of attendees. The authors argued that providing real-time data collection and video delivery via mobile devices are two successful aspects of their system.

TuVista [Bentley 2009] aimed to support the mobile consumption of near-to-live content (from an avg. 15 min till an avg. 30 sec) related to an event. In TuVista, one or more professional video editors monitor a preview of available video streams, delivered from static cameras in the stadium. In turn, the editors prepare so-called multimedia bundles that consist of pre-selected clips from multiple angles and added links to related content such as scored goals or photos and videos previously captured by spectators. Spectators in the stadium can then access these bundles through the in-stadium wireless network (cf. figure 2.2).

TuVista was used as a probe to understand what additional information spectators want to consume during a match in a stadium. It was found that having an appropriate number of clips from key events quickly available was appreciated by TuVista users. This helped the users catch up on a game from afar or relive a play at the stadium. Although the TuVista system has demonstrated the usefulness of mobile sports clips delivered to fans following the game, the authors neither focused on active engagement nor real-time interaction or communication among spectators. Considering this work, we set out to create a system that would support these to fans interested in following the game in the stadium.

In sum, the prior studies showed that there exists a crucial need in designing for fostering awareness within spectators who share the same event and location. The research discussed above already tried to support awareness over the event by sharing sport-related contents on spectators' mobile phone. Among others, it was found that providing real-time data and video delivery via mobile devices are essential factors to be considered. Moreover, it was suggested that systems that aim to foster awareness can enhance feelings of connectedness among not only friends but also strangers. The review of the related work, however, revealed that no system has focused on in situ sharing of user-generated contents created by spectators. These findings guided the design of CoStream presented later in this chapter.

2.1.3 Active Media Creation in Events

In this section, we review previous studies that investigated how and why spectators actively create media (e.g., photos or video recordings) during an event. Moreover, we review systems that supported spectators to create media in an event.

Mäkelä et al. [Mäkelä 2000] investigated the role of mobile phones, particularly digital photos in leisure and communications. They aimed to understand the key motivator to take picture using mobile devices and if this is used for communication instead of text or audio messages. Based on the results derived from a field trial, they found that photos are not only used as memories of special moments of events

Figure 2.3: Users at CityWall [Peltonen 2007].

but also as a tool for creating playful stories and expressing affection.

The study suggested that mobile systems and applications should enable users to create series of photos, edit photos in different ways as directly as possible (due to small screens), and send and receive images with web-based applications. They concluded that image-based communication was considered very practical but was only a completion for the other types of communication, such as functional request, as images are too ambiguous.

Frohlich et al. [Frohlich 2002] explored the role of photo sharing at homes through a field observation and proposed a set of user requirements for future photo-sharing technologies. The results of the field observation revealed that although photographs purportedly are taken to capture memories for personal reference in future, showing off photos taken during an event is a way of sharing experiences with others who were not in the event.

They found that informal communication with images is remarkably common as an enjoyable activity that support deepen personal and community relationships. Moreover, this would allow remote people to quickly exchange viewpoints and explain live event experiences in their domestic surroundings.

Peltonen et al. [Peltonen 2007] proposed an extension to large-scale event participation by displaying user-generated mobile media on an interactive public display. The display, called CityWall, was situated in a city center to show information related to ongoing events taking place around the city (cf. figure 2.3). To be appropriate for a public space, their system supported multitouch and multiuser installation.

They evaluated their system in the field during a music and a samba carnival. They found that users are more present at events through the use of mobile cameras. Moreover, event experiences were relived and wrapped up in a fun way when users browsed through the captured videos and photos of the events afterward together.

Figure 2.4: BannerBattle: The equalizer shows the soundscape and indicates that the fans are singing louder [Bentley 2009].

Figure 2.5: Cheering-meter: The displays are showing the value obtained by Cheering-meter during the concert.[Barkhuus 2008]

This study showed us that passive observers of an event can easily become active spectators if creating and sharing their own content can be viewed as an integrated part of the experience rather than something extra and awkward.

Barkhuus [Barkhuus 2008] developed a system, namely Cheering-meter, that enables interaction between performers and audience. This system distinguished different levels of audience cheering and visualized it on large displays as it is shown in figure 2.5. They utilized the notion of reward applause to engage the audience actively in performances. Cheering-meter is used at eight different rap competitions to enable the investigation of audience-performer interaction. The results revealed that although using technology augmentation with crowds can be very challenging, it provides new ways of interaction that increase the level and sense of participation among the audience. They argued that their system was successful in encouraging active engagement as it used already present behavior, the cheering, for interaction. We therefore consider being immediate and lightweight as important factors for the design of CoStream interaction techniques.

Ludvigsen and Veerasawmy [Ludvigsen 2010] designed a prototype, namely BannerBattle, that displays the same network-connected contents on two different big displays. Each display is eight meters long and is located in front of each fan groups to augment the collective activities of each group such as cheering to support their team. The prototype system is shown in figure 2.4.

To understand how to design technologies for active spectator experiences, they conducted an experiment during several football events. They found that while spectators attend the sporting event instead of watching the game at home, they mainly use this banner to gain a deep understanding of the event itself (i.e., statistics of the match or the performance of the individual players) rather than engaging in the atmosphere. Their findings revealed that the design of technologies to motivate the spectator's engagement should consider the aspect of the sport, the aspect of the event, and the social aspect for active spectating at sporting events.

In summary, the prior work discussed above showed that there is a growing trend and a strong desire for creating and sharing user-generated media and active participation among spectators to go beyond the passive consumption paradigm. Later in this chapter, we show how our work leveraged this trend in the design of CoStream to support coexperiencing the events that are surrounded with social atmosphere.

2.1.4 Mobile Sharing of Multimedia during Events

In this section, we review the previous research that focused on mobile sharing of user-generated multimedia contents between spectators participating in the *same event* to enrich event experiences.

TrottingPal [Nilsson 2004] was a mobile web application system for enabling sport spectators to collaborate at a trotting track event. It was designed to support the users to gather, add, and share information about event's competitors while moving around the arena. The collaboration between spectators is based on either free text messages or with help from default templates. TrottingPal was evaluated during two field trials. The authors found that the application motivated a variety of activities between spectators while communicating toward a common goal. Gathering and building up information, negotiating general facts and tips, and making sense of collected information were instances of such interrelated activities.

Counts and Fellheimer [Counts 2004] proposed a photo sharing system, namely Flipper, to enhance social presence and sharing events. Flipper enables users to flip through photos as with a photo flip book and provides a minimal set of features such as commenting on a photo and sharing within a buddy list. They compared

Figure 2.6: Mobile media sharing in a rally competition [Jacucci 2007b]

their system with the current standard sharing method (e-mail) in terms of ease of sharing and the user experience. The study showed that Flipper successfully increases the number of shared photos and promotes the sense of social presence during life events. Authors argued that people clearly liked to have multiple groups with whom they share their media. They also found that despite the fact that photos are enough for communication and sharing events, additional commenting features (e.g., unanimous audio commenting) can be used to increase people's sense of presence.

Jacucci et al. [Jacucci 2007b] investigated asynchronous photo sharing (i.e., taking a photo and sending it to another spectator). They particularly looked into the potential of such sharing practices for large-scale events, where spectators are scattered across different sites and can only partially witness the whole event. They argued that in such events, spectators experience the event together in other ways than just watching [Reeves 2005]. They explored how capturing and then sharing experiences using mobile phones can be a participative practice to enhance the overall experience during a three-day car race and a music festival (cf. Fig. 2.6). They found that asynchronous media sharing has a potential to facilitate on-site reporting to off-site spectators, coordination of group action, and keeping up-to-date with other visitors or spectators.

While asynchronous media sharing is certainly helpful, live and therefore synchronous media sharing during events was also shown that provides more immersive means for social interactions [Sahami Shirazi 2011, Laiola Guimarães 2011b, Chuah 2003, Bulterman 2013].

Liu et al. [Liu 2007] developed the Zync application as a plug-in module for Yahoo Messenger. It is a synchronized media player that enables a social viewing

Figure 2.7: Real-time nonverbal opinion sharing during sport events [Sahami Shirazi 2011].

experience for online videos to support synchronicity and co-presence around media events. The application is studied in a lab with a number of Yahoo employees to obtain insight into how people hold conversations in synchronicity with temporal media. They found that synchronous communication about shared media events can affect user behaviors. This means that the distribution of chat and the behaviors for an event is a potential measure for levels of interest in certain parts of that. In this way, a model based on aggregate user behaviors may be able to explore areas of interests in the media event.

Sahami et al. [Sahami Shirazi 2011] proposed to share live non-verbal opinions using mobile phones while watching a soccer match (cf. figure 2.7). They conducted an uncontrolled user study in the wild during the soccer World Cup 2010 to investigate if this iconic interaction is reasonable to enable sharing reactions to the events in real time. They found that the aggregated user's inputs, which correspond to important moments in the event, can be used to generate a summary of the event. Furthermore, sharing nonverbal opinion was found to be helpful to enhance the sense of connectedness and enjoyment and minimize user distraction when operating the mobile devices.

Guimaraes et al. [Laiola Guimarães 2011b] presented a set of general guidelines for socially-aware video authoring (editing) and sharing systems that are realized in a coherent system called MyVideos. It incorporates a number of automatic , semi-automatic and manual processes that assist creating personal memories of a small-scale event such as school concerts, where performers and the audience belong to the same social circle. It takes into account two novel factors namely, 1) the emotional intensity by recommending (filtering) people and moments that might bring memories to the user and, 2) intimacy by providing users the means to enrich videos for others by including ultra-personal content such as audio, video, and textual comments. The system (and thus guidelines) were evaluated in two long-term users studies in the context of music school performance. It was found that MyVideo supported users to remembering events and improved social connectedness. It also encouraged users to capture more videos during events. Inspired by this work, this chapter considers the case in which audience capture videos live in a local-scope mass event and leverage this as a mean for stimulating social interactions with others, even strangers.

In summary, the aforementioned systems focused on leveraging real-time media sharing (such as instant messaging or non-verbal icons) to enhance the copresence and coexperiencing an event. To the best of our knowledge, the in situ coconstruction of shared experiences through user-generated mobile live video sharing has not yet been fully explored.

Besides the prior studies and system presented above, there also exists a variety of commercially available services focusing on online synchronous user-generated video sharing, which is more related to the main focus of our work. For example, ComVu is a real-time video broadcasting service, which was launched in 2005 to enable live video broadcasting from a smartphone to a public website. Other similar services are Livecast.com, Qik.com, Kyte.tv, Bambuser.com, Flixwagon.com, CollabraCam.com, Stickam.com, Ustream.tv, and most recently color.com. These services focus on supporting remote sharing, overcoming larger physical distances in real time. These services still lack means for embedding the experience into the specific event and neglect crucial information, such as the location of potential video sources (i.e., properly equipped spectators).

2.1.5 Summary

In this section, we analyzed four different areas related to our research presented in this chapter. Section 2.1.1 introduced related work focusing on the practice of sharing user-generated short text messages (microblogs) around live events. They mainly enabled a high-level summary of an event by visualizing microblog posts on timeline- or location-based interfaces. We found that user-generated microblogs, in particular Tweets, have the potential to provide a convergence of important moments and help understand and explore the trends and the story of events. However, these prior studies did not focus on real-time sharing and live communication, which may generates novel shared experiences during events.

State-of-the-art
– Location-aware visualization of user-generated contents eased following an event.
– Obtaining awareness about the event found to be important and crucial.
– Real-time interactions and communication can enrich the overall viewing experience.
– Becoming an active part of the experience is strongly desired by spectators.
– Available video broadcasting services focus only on sharing video contents over distances.
– Previous studies lacks a strong real-time coupling to the event using user-generated videos.

Table 2.1: Main findings of the related work analysis that are considered in the design of CoStream.

In Section 2.1.2, we reviewed prior studies providing additional information for spectators in order to foster awareness over events, where spectators share the same-place event. The findings analysis shows that providing *real-time* interactions and event-related information in situ can enrich the overall viewing experience. Although it is shown that designing for fostering awareness is a crucial need for spectatorship experiences during live events, no system has focused on in situ sharing of user-generated contents created by spectators at events.

We then discussed the related work that supported active engagement in live events presented in Section 2.1.3. Our literature examination in this field of research revealed that because of the importance of the social aspect of events, becoming an active part of the experience is strongly desired. It was also found that passive observers of an event can easily become active spectators if creating and sharing their contents can be viewed as an integrated part of the experience. This behavior was particularly observed in events that involve a high social atmosphere such as sporting matches in stadium or arenas.

We finally presented an overview of studies, systems, and commercial online services focusing on live mobile sharing of multimedia to bridge the physical distances and support the "being there" [Hollan 1992]. It was shown that bridging distances is still limited to either asynchronous media sharing (i.e., taking a photo or recording a video snippets and sharing it afterward [Jacucci 2007b]) or sharing live instant messages or non-verbal opinions (e.g., emoticons) [Sahami Shirazi 2013]. Our literature analysis revealed that sharing user-generated videos in real time during events is not sufficiently supported in this stream of research. Furthermore, requirements for the interaction design of such systems are still unexplored. While artistic design guidelines do exist (e.g., scaffolding the creative process of video creation on mobiles [Juhlin 2010]), we believe that live mobile video sharing calls for more time-critical and immediate interactions than artistic camera handling. Our work goes beyond the prior work by focusing on live sharing of user-generated videos to support events where spectators share the same event and the same location. Through a user-centered design approach, we also highlight requirements and implications which inform the design of future multimedia systems.

Table 2.1 summarizes important findings of the literature examination discussed above. In the following section, we present the CoStream design process and its interface concepts.

2.2 CoStream

In this section, we present CoStream and its design process. As mentioned above, it is a set of interaction concepts that are coherently implemented as a mobile application to address the in situ challenges of events' spectators. Before describing CoStream design, we recall the challenges through the following scenario:

Alice and her parents decided to go to a soccer match. They bought tickets for the main aisle, since these seats are perfect for maintaining an overview during the match (cf. figure 2.8(a)). However, a detailed view (e.g., on the opposing team's goal), is only available to those with tickets in other aisles closer to the goal. Luckily, Alice discovers that Bob, a friend of hers, just checked into the stadium on Facebook. She messages him and learns that he is with friends near the opposing team's goal. Unfortunately, they cannot enjoy the match together, but Alice calls Bob through a Skype video call on her phone. Seconds later, their favorite team advances and since Bob is close by, he streams the scene to Alice and friends (cf. figure 2.8(b), who can now witness their team strike. They all cheer together by streaming videos of themselves in both directions.

As outlined in the scenario, widespread technologies already provide a certain degree of support, yet emphasize bidirectional video streaming (cf. Skype). We argue that sharing is not only restricted to friends, but also, basically any user-generated video during an event can be harnessed to enhance the event experience. These observations led us to the following research questions that serve as the foundation of CoStream:

(a) (b)

Figure 2.8: Scene from the scenario: main aisle in a soccer stadium. Participants are restricted to their aisle and thus to this very point of view (a). Bob is recording the cheering team after a goal was scored (b).

- How can mobile live media (particularly video) sharing support the in situ experiences and social interactions?

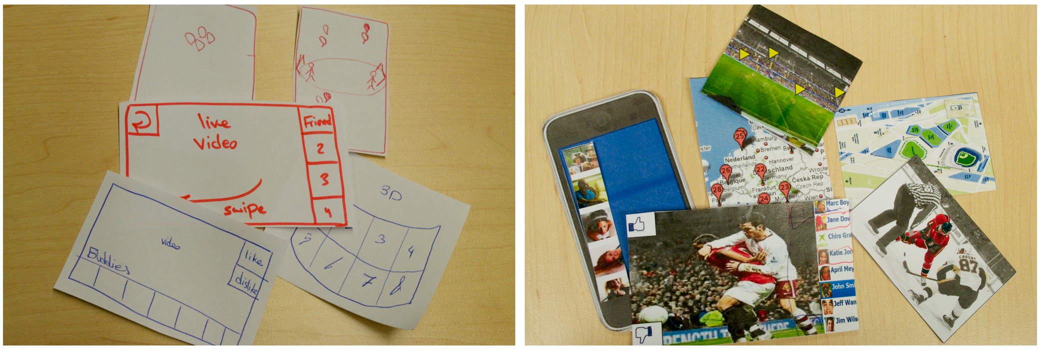
- What are the design requirements for a mobile system supporting this?

- How will it affect the overall event experience?

In the following section, we first elicit interface requirements of CoStream, derived from several focus group studies with real users. We then present CoStream, a novel mobile live video sharing system and its interface concepts.

2.2.1 Requirements Establishment

To empirically establish the interface requirements for the design of CoStream, we conducted three focus group sessions. Each session lasted two hours. Both data gathering and analysis were performed iteratively. We recruited seven participants per session; twenty one in total and different for all sessions. They were between 22 and 34 years old. Each focus group was equally comprised of (1) potential end users, such as passionate soccer fans and two interaction design researchers, who had been working at the intersection of HCI and Multimedia for around four years in average. All were exposed to mobile video creation before, mainly to capture precious moments, while some of them during sports events particularly.



(a) (b)

Figure 2.9: Paper prototypes. Some paper prototypes, resulting from the first session (a). Refined paper prototypes with printed interface elements used in the last session (b).

2.2.1.1 Data Gathering and Analysis

Discussions during the sessions had a brainstorming character, but participants were also involved in creating paper prototypes of their design suggestions. We chose a qualitative data gathering and analysis methodology, which we performed iteratively per session.

In the first session, In the first session, the participants were only introduced to both scenario and research questions. At this point, no prototypical interface was presented. Paper prototypes, generated in the first session (cf. figure 2.9 (a)) were then used as input for the second session. The objective there was to refine and discuss the interface concepts in detail. The refined paper prototypes were the basis for paper mock-ups with printed interface elements (cf. figure 2.9 (b)), which in turn were discussed in the last session.

In addition to paper prototyping, we used video-recording and photo documentation for data gathering. Both data gathering and analysis were performed iteratively. After each session we transcribed the data, selected salient quotes and coded them using an open and selective coding approach [Strauss 1998]. Thus, the analysis results of each session directly impacted the subsequent session.

2.2.1.2 Results and Requirements

Several themes emerged from our analysis of qualitative data that are compiled as four main design requirements. These are discussed in the following.

R1. Support efficient overview and awareness

Participants mentioned they *"want to see who is in the stadium (e.g., friends) and whether a spectator is recording something or not."* In particular, the participants stressed the importance of the efficient access to this information, since they *"do not want to spend too much time looking around for streams."* Indicating the orientation of the spectator was also considered important, since the participants *"want to know whether a spectator is filming in the direction they are interested in."*

R2. Support proper social interaction

Most participants found value in being connected to other spectators in situ. One participant explained she appreciated to *"stream a video selfie of her grimaces to other in stadium."* Several participants discussed that commercially available technologies already provide a certain degree of support, yet emphasize bi-directional video streaming (cf. Skype). As to the social experience, they basically require a priori known user. They argued that sharing is not only restricted to friends, but also, basically any user generated video during an event can be harnessed for experience enhancement.

R3. Support active engagement

While the participants generally liked the idea of being able to connect to friends close-by through video, they mentioned that they would want to *"actively poll other users to stream from a certain perspective"* for them. Moreover, inviting other users to their own stream was considered important, as well as feedback while streaming, for example one participant commented, *"something comparable to the like button in Facebook; it should be easily understandable and just communicate 'hey! I like what I see – keep on streaming."*

R4. Support immediate interaction and reduce visual attention

Throughout our design sessions, the participants underlined the fact that streaming a live situation is highly time-critical, requiring particularly careful interaction support. As one participant put it, *"it must be possible to record moments quickly, without looking at the device."* They imagined this to be ideally as easy as pointing in physical space, *"I just want to point in a certain direction and then see from that very perspective."*

Table 2.2 summarizes the design requirements derived from the iterative design process and the extent to which they are covered in the previous work. In the next section, we describe the interaction concepts of CoStream and how the requirements

Requirements		Supported by prior work?	Contributions of CoStream
R1	Support efficient overview and awareness	◐	CoStream allows spectators to obtain an overview of other spectators and provides in situ awareness through a map and an augmented reality modes.
R2	Support proper social interaction	◐	CoStream supports live sharing of and commenting on user-generated video in stadium.
R3	Support active engagement and social interaction	◐	CoStream allows spectators to actively draw their friends' attention to what they are seeing.
R4	Support immediate interaction and reduce visual attention	◐	CoStream enables to switch between the interface modes via implicit embodied gestures.

Table 2.2: Overview of design requirements. ○and ◐show if the state-of-the-art have covered the requirements to some degree, respectively.

are covered.

2.2.2 Interface Concepts

Based on the design requirements, here we present the interface concepts of CoStream that are divided in three different categories: overview and in situ awareness, watching and streaming, and active engagement. In the following, we first describe the underlying interface concepts and then present each interaction category along with corresponding interaction techniques.

Underlying Interface Concepts

Recalling R4, mobile applications supporting spectators in situ need to provide user interfaces that can be operated eyes-free and inattentively. In the focus group

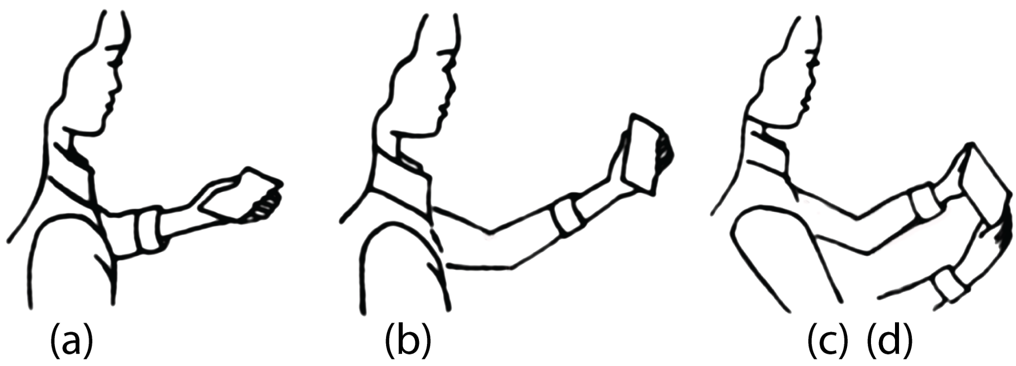


Figure 2.10: The conceptual interaction design of CoStream modes: (a) portrait parallel to the ground, (b) portrait upright, (c) landscape.

studies, users found this as a key for the success of mobile systems to be used while following an event. Moreover, such applications should enable unobtrusive and transparent user experiences that was found as a crucial factor. This means that users should be able to instantly and easily trigger important functions of the application (e.g., streaming). This calls for interaction design that is based on intuitive gesturing rather than interacting with a feature-rich graphical user interface populated with a number of application elements (e.g., buttons).

To meet these application requirements, we leverage various postures (parallel to the ground or upright) and orientations (portrait or landscape) of the mobile device and associate them to the main application features or modes that are naturally more convenient to use. To switch between application modes, users then need to simply turn of the device requiring no visual attention to the application. Today's smartphones are mostly equipped with embedded sensors, such as accelerometer or gyroscope, that can be used to easily recognize device postures and orientations.

CoStream interface takes advantage of three common device postures: portrait held parallel to the ground, portrait held upright, and landscape. These are depicted in figure 2.10.

Portrait parallel to the ground: This posture (cf. figure 2.10 a) is commonly used for an attentive interaction with mobile applications requiring the two hands. In CoStream, turning to this mode exposes application features and UI elements (such as a map UI or showing friend lists and related information). It is designed for situations in which users can visually engage in the application and nothing important is happening in the event (e.g., during breaks or boring moments).

Portrait upright: This posture (cf. figure 2.10 b) supports one-handed interaction. For instance, holding and moving around the device resembling a magnifier glass. We use this posture for turning into an augmented reality view to support in situ discovery and interactions. In this way, mobile device is used as a see-through lens in which user's attention is in the same angle as he is targeting to avoiding attention switching and interruptions.

Landscape: Turning to this orientation (cf. figure 2.10 c) offers an additional layout that was found to provide an immersive experience for particularly watching video contents with a full screen video player and playback controls on smartphones [Sahami Shirazi 2013]. A user's grip on the touchscreen (both thumbs on the screen) can be also used to trigger a special action such as starting streaming, in addition to watching a stream on the display. This orientation can also be used either parallel to the ground or upright. The latter has the same advantage as the portrait upright posture, which supports interactions without any visual disconnection to the event.

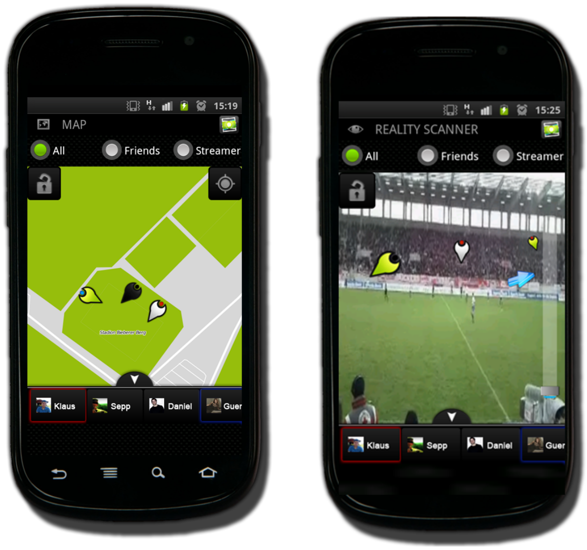
Based on the aforementioned device postures, we designed a number of interaction techniques presented in the next section.

In situ Awareness Techniques

Map visualization: Initially, CoStream provides an overview of the user's current location and of nearby spectators in a "map" preview (cf. figure 2.11 (a)). Users invoke this view by holding the device in the portrait mode parallel to the ground," similar to holding a map in hand. Users can navigate within the map and discover other CoStream users along with their application statuses. Interaction with the map is supported through typical multitouch gestures (i.e., drag-to-pan and pinch-to-zoom).

In situ scanning: This technique provides in situ awareness through an augmented reality view. It is invoked when the device is lifted and held facing the environment (upright) like a see-through display. In this mode, CoStream shows available streams and fellow spectators in the vicinity (cf. figure 2.11 (b)). Using this technique, CoStream fosters immediate interaction (R4), as users are able to just point in a direction to reveal available streams for a particular perspective.

Nearby spectators are visualized in both views using small icons that double as arrows (cf. figure 2.11). The icons show the social relationship to the spectator (friend or stranger) and are oriented according to the direction of the camera (R2, R3). In accordance with the iterative design sessions, this technique aims at conveying the direction a spectator (or rather, her device) is currently looking at.



(a) (b)

Figure 2.11: (a) Map visualization and (b) in situ scanning technique. The arrows are color-coded: green are friends, white are strangers, and black is the user herself. The arrows also contain a dot, designating the current action, whether a user is recording (red), watching (blue) or being passive (black).

Furthermore, icon decorations reveal whether a user is currently recording, watching, or being passive.

Watching and Streaming Techniques

Rotate-to-watch: Once a stream has been located, users can immediately start watching that stream by simply rotating the device into the landscape mode (see figure 2.10 (c)). If multiple streams are available in the considered direction, a thumbnail grid with the latest video frame of each stream is provided. Users can then select a desired perspective by tapping onto the thumbnail. CoStream also supports replaying scenes: playback can be rewound by 30s by tapping onto the circular icon on the left. Tapping again resumes live playback.

Tap-to-stream: To start streaming (cf. figure 2.10 (d)), users can tap and hold down with two fingers anywhere until the camera is ready. This allows users to concentrate their visual attention on the event and just use the device to point at an important scene and immediately start streaming. Tapping again with two fingers ends streaming and allows for an efficient mode switch. If the user is already watching a stream, the playback will be continued and the camera is shown in a picture-in-picture mode (cf. figure 2.12).

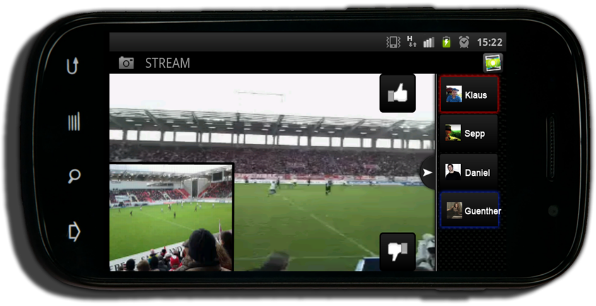


Figure 2.12: A user watches a stream in the center view while he simultaneously streams for others (displayed in the bottom right corner, picture-in-picture).

Active Engagement Techniques

Push-and-pull: CoStream allows users to actively draw their friends' attention to what they are doing (R3): for this purpose, the interface provides an overview of friends in the vicinity in a sidebar for both watch and stream mode. By tapping onto the current video and dragging it to a friend (push mode), the friend is invited to watch the same stream as the user. By tapping onto a friend's icon and dragging it into the video screen (pull mode), the user switches to the same video the friend is currently watching or streaming.

Vibrofy: In addition to visual notifications, CoStream signalizes invitations through vibration. This allows for notifying users in the noisy and laud environment of a stadium.

Rating: The interface also provides a like and a dislike button, allowing users to rate streams (R2). The number of likes/dislikes and the current amount of viewers is shown in the top right corner (cf. figure 2.12). This can potentially provide some hints to select a stream that captures a scene from a better perspective.

Table 2.3 summarizes the interaction concepts of the CoStream interface.

2.3 Field Study 1

We conducted a field study to explore (1) how CoStream is actually used in real-world settings, (2) whether it supports the insitu coconstruction of shared experiences during events, and (3) how it affects social interactions. Furthermore, we

In situ Awareness Techniques		
Name	Purpose	Description
Map Visualization	Support location awareness, looking direction and activity	CoStream provides a map view of the current location and nearby spectators.
In situ Scanning	Support awareness using see-through display	CoStream shows available streams and fellow spectators in the vicinity.
Watching and Streaming Techniques		
Name	Purpose	Description
Rotate-to-Watch	Support immediate video watching	Users can immediately start watching the streams located on map or augmented views by simply rotating the device.
Tap-to-Stream	Support immediate video streaming	Users can tap and hold down with two fingers on the mobile screen to immediately start streaming.
Active Engagement Techniques		
Name	Purpose	Description
Push-and-Pull	Support active and social engagement	Users can drag a video to a friend to invite her to watch. Users can drag a friend's icon into the video screen to invite.
Vibrofy	Support social interaction	CoStream signalizes invitations to watch or to stream a video through vibration.
Rating	Support rating the user-generated video	CoStream provides (dis)like buttons, allowing for rating streams.

Table 2.3: Summary of the interaction techniques proposed by CoStream.

aimed to get insights into CoStream's usability and the overall user experience. In the following, we first describe the study design and the employed methodology followed by results and discussions.

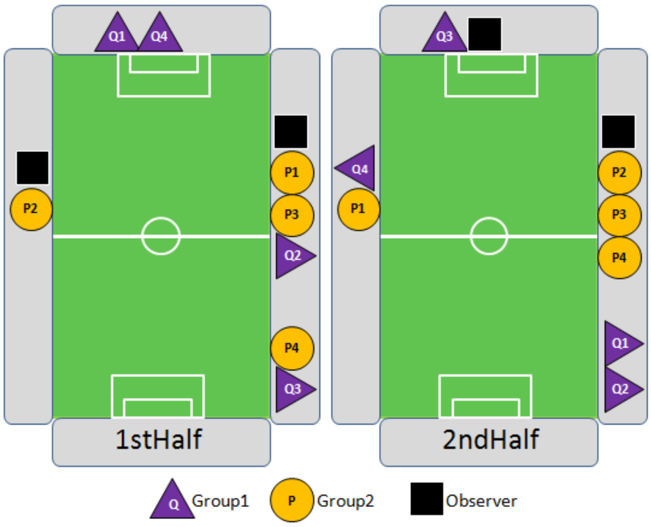


Figure 2.13: Participant locations during the first study. Unfortunately, the south aisle was under construction and no participants could be seated there.

2.3.1 Design and Methodology

The study took place during a mid-scale soccer match with 8190 spectators in Germany. We used a snowball sampling technique to recruit 2 groups of 4 friends (P1-P4 and Q1-Q4; 7m, 1f; avg. 25 years). The groups did not know each other. All of them but two were regular attendees of soccer matches. The two, however, were regular attendees of ice hockey and basketball matches. Four participants (2 in each group) confirmed to use live video sharing services such as Ustream.tv. The participants were introduced to CoStream upfront. The session took about 4 hours. The participants were paid the entrance fee and a mobile data plan.

Apparatus

CoStream is implemented as a client-server architecture. The mobile clients are Android-based and use RTP streaming. The server uses VLC to distribute the videos via HTTP. Thus we achieve a delay of less than 1s over HSUPA. All videos are stored on the server-side. CoStream uses Facebook to manage social relationships between users.

Data gathering

We used interaction logs (video and usage), semi-structured interviews (before, after the match) and observation. Two of the authors engaged with the participants as participant-observers but did not use CoStream at all throughout the study. Figure

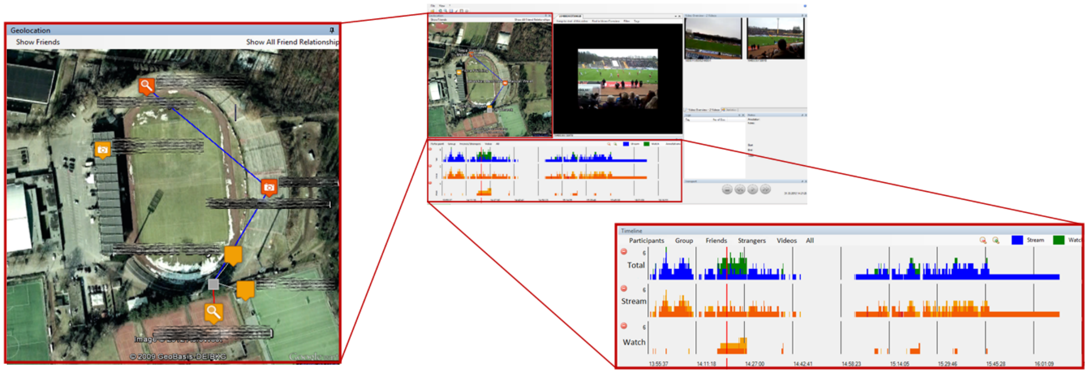


Figure 2.14: The analysis tool has three main views: Thumbnail, Location, and Timeline view with annotations.

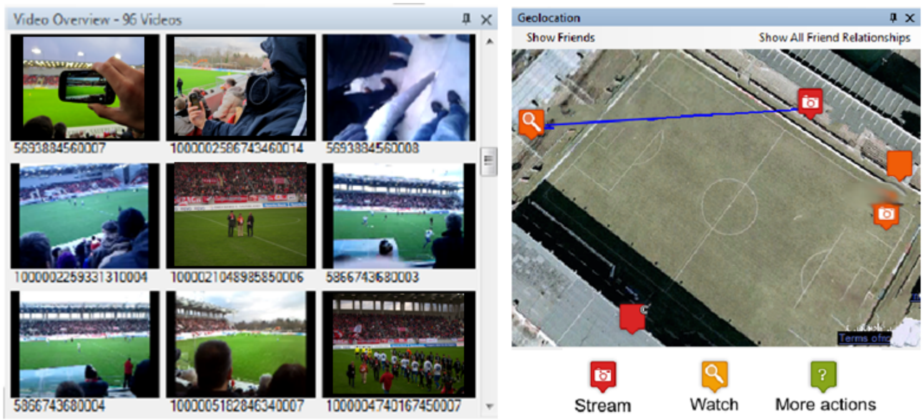
2.13 shows the participants' and observers' location during the two halves of both matches. Neither of them knew their actual location upfront. This arrangement enabled us to observe the behavior of the participants between co-located friends, strangers and distributed individuals during the match.

Data analysis

To analyze the time-coded interaction logs, we designed and developed *LiveVidAnalyzer*: an analyzer tool that visualizes different perspectives of multimodal and time-coded data. It supports the analysis of the recorded videos of the participants with respect to geo-location, social relationship, and sharing behavior. The tool enables advance navigation through different perspectives of same data, such as videos and informs its relation to the geo-location, social relationship, and sharing behavior of video streamers. Furthermore all data is synchronized and aligned with respect to a unique timeline.

Although there exists a variety of tools for this purpose [Fouse 2011], [Hagedorn 2008], [Wittenburg 2006], these tools do not support extensible presentation and navigation of multimodal time-based data from different sources (in our case the recorded videos and related interactions). Furthermore, they do not support the visualization of social relationships between participants and their interaction respectively, which is essential for answering our research questions mentioned above in Section 2.3.

Our analysis tool has three main views (cf. figure. 2.14). Multiple video sources can be visualized on a Thumbnail view (cf. figure. 2.15 (a)). The geographic position of the participants is displayed with a pin on a map on Location view (cf.



(a) (b)

Figure 2.15: (a) Thumbnail view shows an overview over the recorded videos at the selected moment in time. (b) Location view is where the geographic position and the relationship of the participants are displayed on a map. The color coding is used to show the social relationship as a group of friends has the same color. Icons represents how currently each users is doing.

figure 2.15 (b)). The social relationship is visualized through the color of the pin: pins with the same color are friends (i.e., a group during the study). Furthermore, the icon of the pin reflects the current state, whether the respective participant was streaming, watching or passive. The Timeline view shows histograms of the participants' sharing behavior (streaming, watching and a combined view on top). Furthermore, it allows for coding of discrete moments in time or regions.

The LiveVidAnalyzer helped us to not only globally visualize the video streamed and watched during the field study but also enabled annotating, coding, and filtering the data based on different criteria interesting for our investigation purposes. For example, we tagged different parts of the videos corresponding to the important moments of the match and analyzed the participants behavior (cf. figure 2.24). Interviews and observations were transcribed and analyzed using an open coding approach [Strauss 1998].

In the following, we present the findings from our first field study. The coding process yielded various categories, depending on how CoStream actually be used, how it fostered awareness and event context and how it supports the viewing experience and communications in situ.

Figure 2.16: Timeline view visualizes the interaction logs of discrete moments in time. Here color coding is used for participants' sharing behavior, such as streaming, watching, and the accumulating of both on top.

2.3.2 Results

The analysis yielded four categories. We present the results within these below.

Production vs. Consumption

The participants recorded a total of 96 videos, which we classified into 4 categories: 45 streams recorded the match, 26 showed communications, 16 the surroundings (such as side-events happening at the aisles) and 9 were recorded outside the stadium. A total of 43 streams out of the recorded 96 were actually watched. These 43 streams were accessed 85 times, 56,47% from friends; the rest from strangers. Thus the participants focused more on producing, than on consuming.

Figure 2.24 shows the combined histograms for both streaming and watching behavior. Additionally, the graph shows important moments in the match as reported by an after match summary of the match by a local newspaper. These moments correspond to peaks in the histogram. Interestingly, the CoStream usage drops after every peak. Furthermore, CoStream has been used nearly throughout the whole match. While the novelty factor of CoStream probably accounts for a high application usage, the results from the interview provide evidence as to why participants streamed in certain situations. They commented that they either anticipated an exciting situation in the match or wanted to show their view to their friends in the stadium.

Occasionally, we observed that participants did not stop streaming. P3 explained that "streaming generated the pressure to continue streaming for others, even strangers", since he did not want to "annoy the viewers". This also led P3 to put his phone in his chest pocket (see figure 2.17), since he was tired of hold-

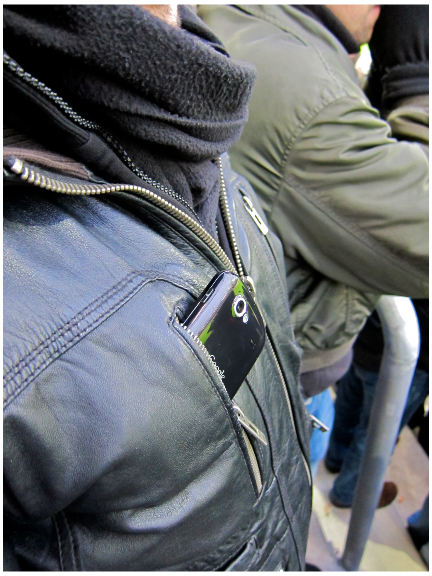


Figure 2.17: P3 stored his smartphone in his jacket pocket to continue recording without holding it.

ing it. We also observed that participants demanded to know whether others liked their stream, since it is *"rewarding", and "an incentive"* (Q1); P1 commented further, *"when I share what I'm recording, I need to know if any one follows my stream."*

Complementing spatial awareness and event context

At the beginning of the study, we observed participants to first use CoStream in order to get an overview over nearby friends. P1 commented in the post-match interviews, *"Once I knew where my friends were, they became a shortcut to their location. I pointed toward my friends' direction whenever I wanted to see the match from their perspective. [...] I had the feeling my friends were in reach."*

The participants repeatedly stressed that peeking at the other perspectives through the live videos fosters awareness for things happening in other parts of the stadium. Q2 commented, *"You get so much more context [...] because I didn't know why spectators standing at the other side of the stadium are shouting angrily."* Moreover, P3 stated, *"The application helps me to know what is going on, on one side of the stadium at the same time of when I'm standing in an opposite side. I could stay connected to the match, even if I left the game for a while to grab something to eat."*

Extending experiences beyond the match

Figure 2.18: P1 asked P2 and P3 to stream. They instantly recorded themselves and cheered with their friend.

The participants commented that they typically switched to a stream when they could not get a good view and hoped for a better perspective. Apart from that, the participants also used CoStream for recordings not directly focused on the match. This is reflected in the histogram, which shows several aggregated peaks, not corresponding to the important moments of the match. On the one hand, the participants used CoStream for sharing side-events, such as singing bystanders. On the other hand, they engaged in playful elements when the match was boring; the situation shown in figure 2.19 took place during the last 7 minutes of the match (cf. figure 2.24), where there was literally nothing exciting happening in the match.

Social interaction and communication

CoStream was often used for communication between friends (cf. figure 2.18) and also to communicate with viewers of a participant's own stream, (i.e., with strangers in particular). For instance when P2 started streaming, he always recorded himself from time to time and waved; he commented, *f"I felt that I had to show that I am recording the stream.*" Participants considered streaming *"something public"* (Q2), since *"everybody is also present in the same event*" (Q3). Thus the participants did not care whether a stranger or a friend watched their stream, since *"even bystanders can see what I record*" (P4).

In the interviews, the participants noted that both push and pull features allow them to immediately interact with others, as Q3 stated, *"inviting friends to a stream is absolutely important for immediately communicating an incident, be it event-related or not, among CoStream users.*" The participants also polled others to stream their surroundings, as Q4 commented, *"I want to see my friends and the*

Figure 2.19: P3 and P4 engaged playfully because the *"match was boring at that moment"*. P4 recorded P3 while he was watching P4's stream, generating a chained recording.

reactions of the fans around them". However, participants also struggled to focus their attention to the event while interacting with CoStream, as P3 described it, *"at one point, I had the feeling that I did not participate in the event.*" We even observed P1 missing the first goal.

Interface usability

Although the participants generally appreciated using CoStream, there were some concerns with respect to the application user interface and interaction techniques. While they stressed the importance of the conceptual subdivision into the four interaction modes, they mainly criticized that they had to switch between the modes via implicit embodied gestures, such as turning the device. They argued that for example turning the device seems appropriate in controlled environments, in a stadium however, other interactions such as cheering with the device in hand accidentally lead to a mode switch. Moreover, they demanded a dedicated recording button instead of the two-finger touch gesture to immediately trigger streaming.

2.3.3 Interim Discussion

The observed phenomena provide evidence that CoStream supports the in situ co-construction of shared experiences in three different ways. First, CoStream enriches *social and spatial awareness*: the participants built a cognitive map of the stadium with their friends being landmarks, therefore serving as quick access shortcuts to different perspectives. This in turn also helped them to overcome social and physical restrictions, since they felt near to their friends. Second, CoStream encourages *active spectatorship*: the participants engaged with CoStream throughout the match to (1) record and also watch other streams and (2) to point their friends' attention to interesting streams they were either watching or recording. The latter also underlines that CoStream enriches *social interactions*: apart from sharing streams, the participants communicated with friends over distance through video or even with the whole audience of their stream (as in the case of P2). Thus, live user-generated video sharing concept during in-situ events has potential to enhance and complement the overall event experience: the participants did this deliberately either by polling others to stream their perspective or by just peeking at other perspectives, therefore gaining a richer context.

In addition to these benefits, study I also revealed a tension between the conventional physical experience of the event and the CoStream-based digital experience of the event. We could not classify this tension as either positive or negative. The more users connected to other participants through CoStream, the more they were distracted from the physical experience, and vice versa. This is also underlined by the results from the histogram analysis. On the one hand, important moments match peaks in the histogram. This shows that participants chose the event as a topic for their recordings, therefore connecting to the event through CoStream. On the other hand, we observed a drop in the CoStream usage right after each of those moments. In the post-match interviews, the majority of the participants stated that in such moments, reactions from the audience focused their attention away from the device back to the immediate perception. However, a contrary effect could be observed when surroundings and match became less interesting: participants intentionally "disconnected" from the actual event in such cases, as described in the scene of figure 2.19.

The results from our first field study are summarized in . In summary, the novel digital experience with CoStream 'competes' with the real-world experience. After study I, it was still unclear whether this tension had an impact on the overall event experience. We decided to conduct a second study to examine the tension in more detail. Additionally, we refined CoStream's user interface in accordance to the usability results.

2.4 Field Study 2

The results of first field study discussed above introduced an unexplored tension between the conventional physical experience of the event and the CoStream-based

General Findings
– CoStream enriched the social and spatial awareness and help to overcome social and physical restrictions.
– CoStream proposed a cognitive map of the stadium with the user's friends being landmarks.
– CoStream served as a shortcut to access different perspectives.
– CoStream made users to be more engaged throughout the match.
– CoStream revealed a tension between the physical experience and the CoStream-based digital experience of the event.
– The novel digital experience offered by CoStream was very well-received.

Table 2.4: Summary of the main findings from first field study using CoStream.

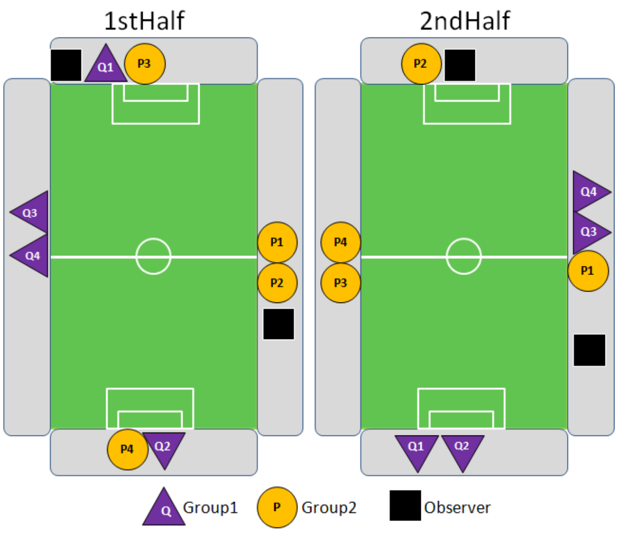


Figure 2.20: Participant locations during the second study.

digital experience of the event. In the present chapter, we conducted a second field study during a mid-scale soccer match with 4500 spectators to address the aforementioned open research question in detail.

2.4.1 Design and Methodology

The study design was analogous to study 1 presented in section 2.3. Here, we recruited another 2 groups of 4 friends (P1-P4 and Q1-Q4; 7m, 1f; avg. 27 years), who did not participate in the first study, using a snowball sampling technique. All of them were regular attendees of soccer matches. They were introduced to CoStream

Figure 2.21: (a) Map view, (b) In situ awareness. Revisited views, left: map view, showing spectator's Facebook avatars instead of the original color coding. All views now have a dedicated record button. Tapping it switches to stream mode, explicitly. Right: stream mode, notification area is highlighted.

upfront in a hands-on session and were paid the entrance fee and a mobile data plan. Overall, the session lasted about 4.5 hours.

Apparatus refinements

The results of the first field study showed that the concrete implementation of switching between CoStream's four conceptual modes was inappropriate and not explicit enough. To address this issue, we introduced a dedicated record button to every view, once being tapped, switches to the stream mode explicitly (cf. figure 2.11 (left)). Furthermore, the results from our study showed that active spectators, who are currently streaming a video, require awareness about if other users are watching their very stream. We therefore added a notification area at the top of the interface, showing the current viewers (cf. figure 2.21 (right)).

We employed the same data gathering and analysis methodologies as in our first study. Figure 2.20 shows the participants' and observers' location during the two halves.

2.4.2 Results

The participants recorded a total of 106 videos during the match. Similar to the results of the first study, the videos' contents can be classified into three different categories: 54 videos recorded the moments related to the match, 47 showed social communication and 12 the surroundings. Note that several videos are classified in more than one category. A total of 58 streams out of the recorded 106 were actually watched. These 43 streams were totally accessed 80 times, 53.33% from friends. Figure 2.24 shows the accumulated usage histogram for both streaming and watching behavior, which indicates that the important moments of the event such as goal chances or fouls again match the peaks. This is in line with our qualitative findings, as Q3 stated, *"I want to stream and share the match-winning scenes"* and confirms the results from the first study. We present the results from the second study below.

Figure 2.22: P1 put his device into his chest pocket, continued streaming and cheered for his team.

Required attention while recording

Throughout the study, we observed that participants easily streamed video for a long period of time, while preserving their attention to the actual event (cf. figure 2.23). P2 commented, *"I was continuously streaming video during the second half, since the players were frequently approaching the goal and I hoped to stream a strike."* P1 mentioned, *"I put the device into my chest pocket while streaming and simultaneously clapped hands to applaud my team"* (cf. figure 2.22). She also added, *"Holding the mobile device for long can be very tiring and I need something to reduce this fatigue and physical effort if chest pocket is not an option."*

Figure 2.23: P2 is streaming and not focusing on the display, but on the actual event.

Coupling between physical and digital experiences

Our observations revealed watching video (as a digital experience) required more attention to the device. The participants repeatedly stressed that it is important for the stream to be live, since *"the stream itself fosters awareness over the current situation in the stadium"* (P3). In line with this is Q3 commenting, *"It was great that the stream was live. So thus what I watched on the device, matched what I heard from the atmosphere and other spectators around me."* However, in the post-match interviews, the majority of the participants stated that reactions from the audience focus their attention away from the device back to the match (as a physical experience). P4 noted that *"when the spectators are cheering, my attention draws back into the match."* Occasionally, we observed the participants use CoStream's replay functionality to replay certain scenes. Q1 said, *"during a match in stadium without CoStream, there is no chance to replay scenes, particularly those of other spectators around."* However, P4 added, *"I imagine that if a match were really intense with lot of action, I would not use the replay function, to not miss anything."*

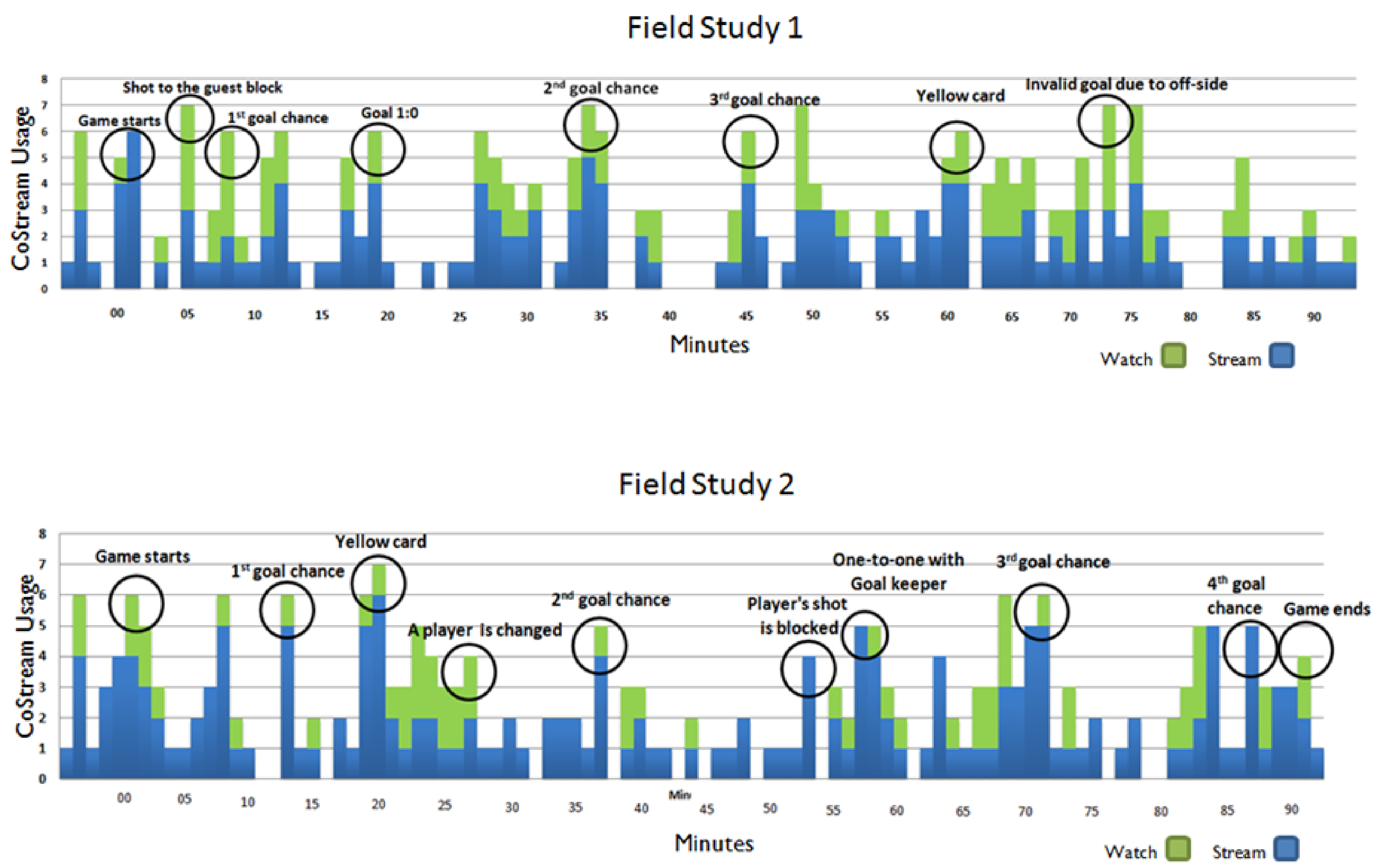


Figure 2.24: Usage histogram of the CoStream application during the first and second field study, peaks correspond to the important incidents of the matches.

General Findings
– CoStream contributed to the event through a strong real-time coupling between physical and digital experiences.
– The tension between CoStream and the actual event could be characterized as an interplay of both experiences.
– CoStream enhanced and intertwines with the overall event.
– The designed user interface enabled the fluid transition between CoStream and event.

Table 2.5: Summary of the main findings from the second field study of CoStream.

Synchronous communication
The participants did not want to use any additional communication such as audio, as P2 stated, *"A soccer match is too loud. This makes audio communication almost impossible. And texting is certainly not an option, it requires too much attention"*. Q4 elaborated on this *"and if I sent my friends a message related to the match, they might not read it immediately and later on, wouldn't understand it"*.

2.4.3 Final Discussion

As outlined above, the first study revealed a tension between the actual event and the application usage. As a consequence, spectators might become disconnected from the event and miss important scenes. The results from the second study highlight two important phenomena: (1) CoStream contributes to the event through a strong real-time coupling between physical and digital experiences and therefore (2) the tension between CoStream and the actual event can be characterized as an interplay of both experiences.

We propose to conceptualize this interplay as a focus+context [Baudisch 2001] approach to experiencing events with ubiquitous live multimedia sharing. While the participants were watching a live stream, CoStream was their focus. The actual context with respect to the event was preserved through both listening to the atmosphere in the stadium (e.g., cheering sounds) and peripheral vision (e.g., immediate reactions of bystanders). Vice versa, while streaming, their attention switched away from CoStream (cf. Figures 2.22 and 2.23), the event became the actual focus; CoStream was still in the peripheral context (e.g., visually or vibrotactile). Moreover, both event and CoStream provide means to support the fluid transition between focus and context. On the one hand, the atmosphere can draw a user's attention to the event. On the other hand, CoStream supports interactions such as pushing and pulling video streams to users, therefore demanding their attention. Our studies show that CoStream enhances and intertwines with the overall event. Key to this is that:

- both physical and digital experiences concern the very same event.
- the system provides a strong real-time coupling to the event, the spectators' locations and their social relation.
- the UI enables the fluid transition between focus and context through (a) providing efficient overview and awareness, supporting (b) active engagement, (c) immediate interaction and (d) reducing visual attention.

Otherwise physical and digital experiences will be decoupled, leading to disconnections (sensu Turkle [Turkle 2011]).

2.5 Conclusion

In this chapter, we address challenges that spectators are faced with when participating local-scope mass events (e.g., sporting matches). Typically, the venue of such events is designed in a way that every spectators has and is assigned to a seat to view the proceedings. Although, spectators share the very same location, the viewing perspective and social interactions with other spectators are somewhat limited to their vicinities and adjoining spectators. To extend these limitations, we propose live sharing of user-generated video contents that is an increasingly popular form of social interaction particularly during live events.

Therefore, in the present chapter, we investigated the potentials for mobile video sharing as a means for providing additional event-related information to spectators in a stadium-based event. Prior research has shown that sharing user-generated contents have a great potential to provide a convergence of important moments and help understanding the trends and the story of events. Moreover, it is shown that providing real-time interactions and event-related information in situ can foster awareness and enrich the overall viewing experience. Thus, we contributed a set of novel interaction concepts, coined as CoStream, for supporting in situ sharing of user-generated videos during events.

CoStream was developed through an iterative design process, starting with several focus group studies to establish its interface requirements. Based on that, we propose interaction concepts and techniques that aim to support efficient streaming and watching activities, obtaining efficient overview and awareness, and encouraging active engagement in events.

We evaluated our system and its user experience in two field studies during soccer matches. The field trials demonstrate that real time sharing of different perspectives on the same event has the potential to provide fundamentally new experiences of same-place events, such as concerts or stadium sports. We found that CoStream supports the in situ coconstruction of shared experiences in various ways. CoStream helped overcome the inherited challenges (namely limited social interactions and physical restrictions) of events taking place in all-seater venues. We discussed how CoStream enriches social interactions, which increased context, social, and spatial awareness, thus encouraging active spectatorship. We further contributed a set of key design guidelines and implications for the design of future interfaces. We believe that future systems, adhering to the design guidelines of CoStream, will pave the way for new possibilities to coconstruct shared experiences in situ. This can be achieved through the ubiquitous sharing of multimedia intertwined with physical event experiences in real time.

Chapter 3

Supporting Interaction for Living Room Experiences

Contents

In the previous chapter, we focused on enhancing in situ user experiences for spectators during live events. We proposed leveraging mobile live user-generated video sharing to address the decreased viewing and social experiences imposed by the physical restriction of in situ events. While these happenings attract a large number of spectators to directly witness and experience the life atmosphere in situ, an even larger number of remote viewers follow the very specific topic through media coverages in *living room settings*. In addition to the live professional broadcasts, the STV supports the availability of a multitude of various multimedia contents and social remote communication for viewers at homes. Therefore, home viewers can get the most out of the event from a distance. This however raises the user interaction challenge for viewers who are following the event at homes, typically in a lean-back manner relaxing in living rooms with a remote control in one hand.

The current way of interacting with televisions typically requires mediator devices, such as hand-held remote controls or touch-based interfaces on smartphones. While this is a well-established paradigm, it has its downsides. The device itself can be easily out of reach and misplaced or interactions may require a lot of visual attention causing user distraction. The present chapter addresses these issues by proposing a novel interaction style between television and viewers at homes that is based on the *human body*. We explore novel, body-based interaction concepts and investigate how can user input for interactive televisions be redesigned to become more usable and offer a more delightful user experience. We argue that leveraging the human body as an interface for the television has various advantages, such as being omnipresent, deviceless, and eyes-free, and thus can potentially enhance the experience of TV watching activity in living room environments.

To explore how a user's body can be leveraged to mitigate user interaction with television and how designers can benefit from such knowledge, we first need a better understanding about people's use of body in front of the TV. Therefore, we start our investigations by observing people and their interaction styles with TVs in a field study. The study particularly looks at how people spatially situate themselves in front of the TV and how they engage in watching activity. Moreover, it focuses on how people could use their body to perform tasks that are typically carried out through remote controls.

The study findings show that the whole body information, such as pose and orientation, have the potential to support coarse-grained TV interactions to execute for example turning the TV on and off and showing EPG or in general, and program awareness. The fine-grained TV interaction (e.g., channel navigation and selection of items) can benefit from spatial movement specified, in particular through the user's hand. Based on the findings of the user study, we identify and outline a set of requirements that are supported through the design of two novel body-based TV user interfaces, namely *CouchTV* and *PalmRC*.

In CouchTV, we support various course-grained interactions with TV systems that rely entirely on the spatial and postural information of viewers. We exploit these information, such as user's presence, location, orientation, and pose, to design novel interaction techniques supporting (re)engaging in TV watching activity, providing appropriate level of awareness, and displaying supplementary information related to TV programs. The CoachTV interface is evaluated in an initial user feedback session with 12 groups of people.

In contrast to CouchTV that leverages the entire body, PalmRC appropriates only the palm of the hand as a means to enable fine-grained interactions with the TV. It is a novel eyes-free input style for television systems that allows TV viewers to perform spatial interactions with empty hands. The underlying concept of PalmRC is inspired by the sense of proprioception that enables human to sense the relative position of their limbs in the space. Unlike typical TV input modalities, such as remote controls or smartphones, users can operate television through touching the palm of their hands with the other hand index finger in an eyes-free manner. This allows user to map remote controls functionalities to their hand and perform fine-grained interactions, such as navigation in menu using arrow keys that are mapped on appropriate salient regions of the palm. The PalmRC interface concepts were evaluated through a set of user studies. We initially conducted an exploratory study to empirically ground the requirements for designing an eyes-free, palm-based TV remote control. Based on the findings from this study, we evaluated our concept through a controlled experiment to investigate the precision and effectiveness of using the palm as an input surface. At the end, we focused on identifying respective advantages and disadvantages of our concept, the traditional remote controls, and touch-based smartphones in a comparative study.

In summary, the contributions of this chapter are the following:

- The preliminary study to examine how TV viewers can use their body as an interface for TV systems
- Design of two novel body-based TV user interfaces (CouchTV and PalmRC)
- Evaluation of the concepts through a series of user studies
- Implications for future body-based TV interactions

The present chapter is organized as follows: In Section 4.1, we first review the state of the art and provide a background of relevant research on user interfaces leveraging whole-body proxemic and spatial information and also take a look at the prior research on hand-based and imaginary interfaces. Section 3.2.1 presents the preliminary user study and outline a set of requirements for designing novel TV user interfaces that leverage the human body, such as the head orientation and hands.

Drawing upon the results of our preliminary study and literature analysis, we depict two novel body-based user interfaces, CouchTV and PalmRC, that are presented in sections 3.3 and 3.4 along with their evaluations, respectively. This chapter is closed with a concluding discussion regarding the effectiveness, efficiency, user experience, and overall preference of the proposed techniques compared to the most typical existing input modalities and several design implications in Section 3.5.

> **Contribution Statement:** Most of the work presented here has been published in [Dezfuli 2012d, Dezfuli 2012b, Dezfuli 2012c, Dezfuli 2014]. I am the first author on these publications and I have initiated and lead the project. My coauthors have also contributed significantly. Master students, Florian Müller, Manolis Pavlakis and Mürat Özkorkmaz have built and implemented many aspects of the CouchTV and PalmRC systems. My supervisors, Jochen Huber, Mohammadreza Khalilbeigi and Max Mühlhäuser have contributed to the design of the systems and helped in writing the papers.

3.1 Related Work

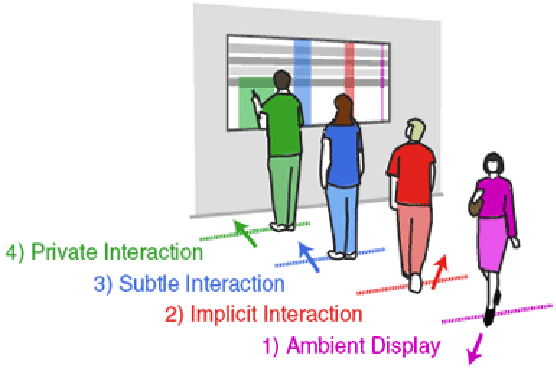
This section provides scientific foundations for the two contributions , namely CouchTV and PalmRC, presented later in this chapter. We sample the related work out of two main areas in the literature including whole-body interaction and hand-based interaction.

3.1.1 Whole-Body Interaction

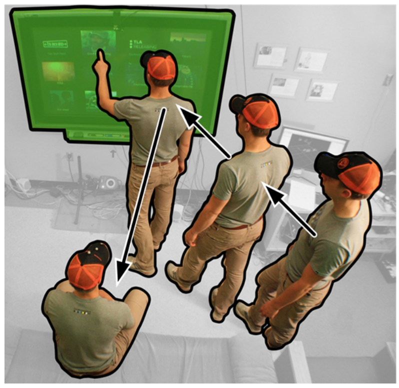
We first review previous systems that leverage human whole body to support coarse-grained interaction with interactive displays. We first focus on body-based proxemic interaction with large surfaces or small display. We then take a look at spatial body-based interactions in the context of television and finally, we review prior work investigating interactions via human vocalization and gestures that required no visual attention while interacting with the TV.

3.1.1.1 Toward Proximity-Aware Surfaces and Displays

In the following, we review related work that use human body proxemics as an interface for interaction with large interactive surfaces and small displays.



(a) Interactive Public Ambient Displays [Vogel 2004].



(b) Proxemic Interaction [Ballendat 2010].

Figure 3.1: Body-based interfaces that leverage user body and postural information to mediate interaction.

Ju et al. [Ju 2008] illustrated that interaction from afar is public and more implicit, while people move towards the surface when they would explicitly interact with more personal information.

Similarly, Vogel et al. [Vogel 2004] explained how a user's contextual body orientation and position can be leveraged for interaction with private information on the large display. They present this concept with the design of an interaction framework for ambient displays which represents information in public, semi public, and private spaces using proxemics (cf. figure 3.7 (a)).

Ballendat et. al. [Ballendat 2010] proposed proxemic interactions in ubiquitous environments based on interpreting spatial relationships of persons, objects, and digital devices (cf. figure 3.7 (b)). They particularly investigated how fine-grained proxemic knowledge (such as user's presence, position and orientation) can be exploited to design interaction techniques with surrounding digital devices.

Several studies also considered how proxemics can influence the way we interact with small displays. Harrison and Dey [Harrison 2008] introduced Lean and Zoom as a proximity-aware interface to control the content magnification. Their system considers a person's proximity (i.e., user's head distance) to a small display such as notebook to magnify the on-screen content for people spending long hours sitting in front of computers. This proposes a reliable way to easily see the digital content with extra or less details based on user preferences and even avoid probable vision-related problem.

Medusa [Annett 2011] is a proxemic-aware multi-touch tabletop in which users

(a) SixthSense [Mistry 2009] (b) Skinput [Harrison 2010]

Figure 3.2: Projection-based on-body interaction.

are locating around the display surface and their touch inputs are associated with them. This system is designed to facilitate multiuser collaborative interactions but does not afford user identification.

Scott et al. [Scott 2003] showed that natural location-based partitioning already occurs while working on tabletops and explained how this can positively impact the collaboration. They illustrated that this approach can ease maintaining workspace awareness, support functionality in the appropriate locality, and have much more or enough table space, which are beneficial for tabletop collaborators.

The prior studies stated above showed the importance of considering user body and postural information for designing future interactive surfaces. In these studies, designers have used the notion of space and proxemics centered on the body to mediate people's interactions with surrounding displays and devices. Most of these works are inspired by the study of Edward Hall [Hall 1990] on human spacing behavior in 1914. Based on his observations on human beings, he identified four bodily distances, intimate, personal, social, and public zones, to represent how humans unconsciously use distance sensing in social situation to demonstrate their relations to other people. While this stream of research definitely helps understanding how the body and postural information of people may help to mediate the communication between human and computers in general, they do not sufficiently describe how this can be leveraged in TV interactions, which is a fundamentally different experience.

3.1.1.2 Toward Body-Aware Television

In this section, we focus on prior studies addressed the role of human body in designing and enhancing TV interaction and TV watching user experiences in living-room settings.

Geerts and Groof [Geerts 2008a] have pointed the necessity of awareness tools for indicating the presence of other users to support remote interaction as sociability design heuristic.

Hobson [Hobson 2008] found that TV viewing can simply incorporate everyday life activities and is not only a spare-time or a leisure activity. She argued that people already developed interesting ways of half-watching television and using TV as an environmental source, such as punctuation of time or just listening to the program as a background noise.

Moreley [Morley 1986] presented a household-centric study to understand individuals in the social context of TV viewing and its relation to everyday domestic life. He argued that not only may ethnicity and gender impact TV watching activity, but also, the TV viewing behavior may differ while watching alone or socially between distinct collocated viewers.

Hawkins et.al. [Hawkins 2005] explored the visual attention of viewers to the TV content based on the length and frequency of looks at TV displays. The author argued that how the attention and the average gaze length can be decreased while watching activity becomes more complex. Similarly, Chorianopoulos [Chorianopoulos 2004], proposed that TV viewers may have different levels of attention to the display, such as one who is fully eager to follow the content.

Wang et al. [Wang 2006] developed game prototypes being aware of a real-time player's face position to enhance the sense of presence and role-playing during the game experience. As a step toward novel context-aware television, they focused on entertaining purposes, where user's spatial information could make the games more engaging.

Overall, these findings are certainly helpful and provide first insights into the TV watching activity and reflect the role of viewers' context (i.e., user's identity and attention). In the present chapter, we go beyond the prior studies and systems by exploring the correlations between viewers' spatial situations and the level of engagement in the watching activity through our preliminary study presented in Section 3.2.1. We investigate more fine-grained TV watching levels of engagement including intermediate ones between fully engaged and total distracted.

3.1.1.3 Eye-Free TV Interactions via Human Vocalization and 3D Gestures

While there is few related work that use body and postural information as an interface for interaction with TV systems, considerable research focused on eyes-free and deviceless interaction by observing and interpreting viewers' vocalization [Igarashi 2001, Brutti 2010], movements and gestures [Freeman 1994, Mäntyjärvi 2004] to interact with TV systems. Among others, speech and 3D mid-air gesture input modalities draw the attention of many researchers as well as TV manufacturers due to supporting eyes-free TV interactions.

Brutti et al. [Brutti 2010] presented a distant-talking interface for the interactive control of a TV set with multi-channel acoustic data collection. Igarashi and Hughes [Igarashi 2001] focused on direct control of interactive television by using nonverbal lowlevel features of voice such as pitch and volume.

Although speech is a natural input modality, its usage is not always socially appropriate. Furthermore, technology may fail to recognize commands in noisy and unpredictable acoustic environments. Besides being inefficient and not wellscalable, it is also not suited for common continuous interactions, such as scrolling a channel list or adjusting the TV volume.

Many studies aimed at addressing these limitations and investigated how viewers can control TVs using 3D hand gestures [Chen 2010]. Freeman and Weismann [Weisz 2007] have investigated how viewers can remotely control a television set by hand gestures without extensive user training and memorization. To do so, they provided visual feedback on the TV screen. This enabled users to move an on-screen pointer coupled to their hand to adjust various graphical controls.

Mäntyjärvi et al. [Mäntyjärvi 2004] explored a possible set of gestures suitable for controlling home appliances such as a TV. They showed that 3D hand gestures lack an easy memorable and universal vocabulary. They reported that mid-air hand gestures are not appropriately recognizable for unpredictable scenes and suffer from scalability issues in group-watching experiences. In addition, their study showed that people find mid-air gestures somewhat uncomfortable and tiring (the fatigue problem [Lenman 2002]). They also criticized the lack of haptic feedback while mid-air gesturing.

The studies in this vein of research have shown that deviceless and eyes-free styles of interaction with TVs are highly desired and avoid distractions caused by other attentive input modalities – such as, remote controls and smartphones. Rapid developments in the field of 3D depth sensing technology (e.g., Microsoft Kinect depth camera), successfully found its way into home for only entertainment purposes. However, the aforementioned drawbacks might explain why speech and 3D gestures as the most advanced eyes-free input modalities are still limited to lab environments and not yet widely deployed in home television environments. Building up what has been proposed in this line of research, our work investigated interaction by leveraging a viewer's hand as a palm-based imaginary user interfaces in an eyes-free manner.

3.1.2 Hand-based Interaction

In this section, we review prior studies on wearable and mobile systems that leverage and augment the surface of the hand and arm as an attentive input system. While not directly relevant, we believe that all these veins of research provide valuable source of inspirations for the design of CouchTV and PalmRC interfaces. In the following, we review related work focused on hand-based wearable input interactions and imaginary interfaces.

3.1.2.1 Wearable Interfaces

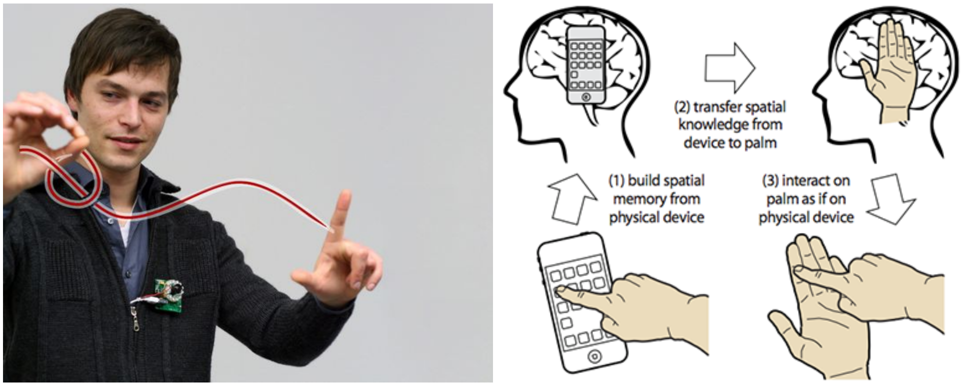
There are various technical approaches to support fine-grained interaction on the surface of the hand.

KITTY [Kuester 2005] is a glove-type input device, that covers parts of the hand with electronic contacts to enable touch event detection. An electric circuit is closed and a signal is generated upon closing of one finger-contact with one thumb-contact. This offers both speed and accuracy with a discrete signal input that is continuously ready and provides an ultra-portable solution for data input into portable computer systems.

SixthSense [Mistry 2009] is a wearable camera-projector unit supporting gestural manipulation of digital artifacts. It augments physical surfaces with digital information and enables users to interact with projected information in mobile contexts (cf. figure 3.2 (a)). While the system is superior to existing systems in terms of weight and size, the system uses color markers as artificial features which are put on a user's fingertips to recognize hand gestures.

Skinput [Harrison 2010] presents a novel approach to recognize finger tap on arms and hands by analyzing mechanical vibrations that propagate through the body (cf. figure 3.2 (b)). The system uses arrays of bio-acoustic sensors, which need to be worn as an armband. Brainy Hand is another example of a wearable interaction device. It is equipped with a color camera, which captures an image of

[0] http://www.microsoft.com/en-us/kinectforwindows/



(a) Imaginary Interfaces [Gustafson 2010] (b) Imaginary Phone [Gustafson 2011]

Figure 3.3: Imaginary interfaces, examples of eyes-free and omnipresent input style.

the user's hand to recognize its movements as input gestures. Since the digital data corresponding to each input gesture is projected as a picture onto the user's palm, it requires a lot of visual attention.

Recently, Harrison et al. [Harrison 2011] presented OmniTouch, a wearable projection-based prototype, enabling multi-touch applications on everyday surfaces including the body. They used depth sensing technology to track a hand and recognize whether a finger has hovered over or touched the hand surface. This work and the proposed touch recognition algorithm inspired the design of PalmRC prototype.

Similar in nature, Armura [Harrison 2012] is an interactive on-body projection system that supports both input and graphical output on a user's arms and hands. The authors explored the design space of arm-driven user interfaces by proposing various synergistic arm gestures and atop of that, developed several interaction techniques and applications. However, the aforementioned research requires either a mediator device or visual attention.

3.1.2.2 Imaginary Interfaces

Since our focus is on TV interaction without any instrumentation on the body, as this is not practical for TV rooms and also can mar the user experience while watching TV, we particularly propose a deviceless approach where the visual attention remains focused on the TV screen. To do this, we draw upon the concept of imaginary user interfaces [Gustafson 2010] that the actual interface elements are not visually projected onto the interactive surface. They are just imagined by the user. Imaginary interfaces are introduced as a new deviceless interaction approach that are based on a human's ability to map the spatial memory to physical surfaces.

In addition, aside from imaginary interfaces , no user interface is displayed on the surface but various sensing approaches are utilized to recognize on-surface interactions. Although, no information is projected on imaginary interfaces, the original concept requires users to look at their hands to define the origin of an imaginary space and attentively point and draw in the resulting physical space.

Building on this work, Gustafson et al. [Gustafson 2011] designed an always available imaginary phone, where users can interact with their cell phone by recalling, mapping and touching different application icons on their hand attentively. This prior work motivated the design of the PalmRC interface. We go beyond what has been proposed in this line of research by investigating interactions that leverage viewer's hand as a palm-based imaginary user interfaces in an *eyes-free* manner.

3.1.3 Summary

The examination of the related works discussed above is summarized in table 3.1. Related to the main contributions of this chapter, we reviewed prior studiesthat that uses the body as user input to interact with displays and interactive surfaces. We also revisited previous works that explore the role of the human body in TV watching as an activity. We learned that TV viewing is blended with everyday life activities and, therefore result in different levels of engagement based on the human spatial situation in front of the television. The prior studies motivated us to further investigate the levels of engagement as the point of departure of this chapter and understand how human body spatial situation can be interpreted to mediate communication with TV watching activity.

Furthermore, the state-of-the-art review reflected that being omnipresent, eyes-free, and deviceless are three key requirements to design user input for televisions. At the end, we discussed that leveraging the part of body, particularly the hands, as a means for TV input can potentially help to ease and enhance TV watching experience.

3.2 Understanding TV Watching Activity

3.2.1 Preliminary Field Study

This study aimed to take a close look at today's people watching behavior in order to gain a better understanding of how TV viewers get engaged in watching activity. Particularly, we designed this study to answer the following questions:

State-of-the-art
– Deviceless, omnipresent, and eyes-free user input can high potentially enhance interaction particularly, the TV watching activity.
– Use of space and proxemics centered on the body can potentially mediate people's interactions with surrounding devices.
– No previous work has focused on leveraging body to enhance TV interactions to create fundamentally different TV watching experience.
– Previous systems mainly leverage and augment the surface of the hand as an *attentive* input system supporting fine-grained interactions.
– Being omnipresent, eyes-free, and deviceless are three key requirements to design user input for televisions.

Table 3.1: Summary of the main findings based on state-of-the-art analysis. This highlights the requirements that are addressed with the contributions of this chapter.

1) How do people spatially situate themselves in front of TV? and how they engage in a TV program?

2) Is there any correlation between their spatial (or postural) situations and the levels of engagement?

3.2.1.1 Method

We observed 15 volunteers (seven female) watching TV individually or collocatedly in their living rooms. The participation was voluntary and no compensation was provided. We recorded and analyzed participants' activities in their living rooms for two weeks in order to be sure that we take part in the day-to-day routine of our participants' households and have well-grounded findings. All were assigned a specific area in their living room for watching TV including a sofa, a small table, and a TV set. Each household watched TV at about ten hours on average (SD: 2.7). At the end of the study, a semi-structured interview was conducted with each participant.

The study resulted in 50 hours of video materials that were coded by two individual coders using an open and selective coding approach [Strauss 1998].

3.2.1.2 Results

Our video analysis revealed participants posed differently in front of the TV. We observed that, as the level of engagement to TV content changes, participants usually change their spatial and postural situations. They choose sitting or lying postures while watching but they also stand before being fully involved in watching, for instance while zapping. P2 states, *"It's not always as easy to relax, when you know that at any moment you may need to look for the remote control to give a command to the TV"*.

Participants frequently were interrupted from watching by their own or because of others. This either caused them to leave the room (e.g., smoking cigarettes) or they were only distracted from watching (e.g., receiving a phone call). Depending on the type of interruptions, they showed different behaviors. Either they started looking for a new program, asking other coviewers about missing scenes, or surfing the web or EPG for more detail about the new TV program starting while having interruptions.

In interviews, participants stressed on the fact that re-engaging with the content is almost difficult and can be time critical. P11 added, *"It can often happen that I do not understand what the actors are talking about once I am not in the room because of any reasons.*" In addition, they frequently mentioned that they require awareness for something special to be shown on the TV and proposed to do this through (e.g., *"visual cues on TV screen"*). P6 said, *" I usually miss my favorite TV shows. It would be great if TV screen could display which of my favorite program is currently running when I pass the TV screen.*"

Levels of Engagement In Watching Activity

Based on the qualitative analysis of the videos, we defined five different levels of engagement for video watching based on presence, location, orientation and posture of TV viewers:

Passive is assigned to people who are present in the TV room but for other reasons rather than watching TV (e.g., to water plants). In this level, either people are not located in front of TV yet or they do not face towards the TV. Although the passive level reflects no or weak connection between people and TV content, detecting this level allows TV to stay inactivated and avoid inappropriate reactions to the person's presence.

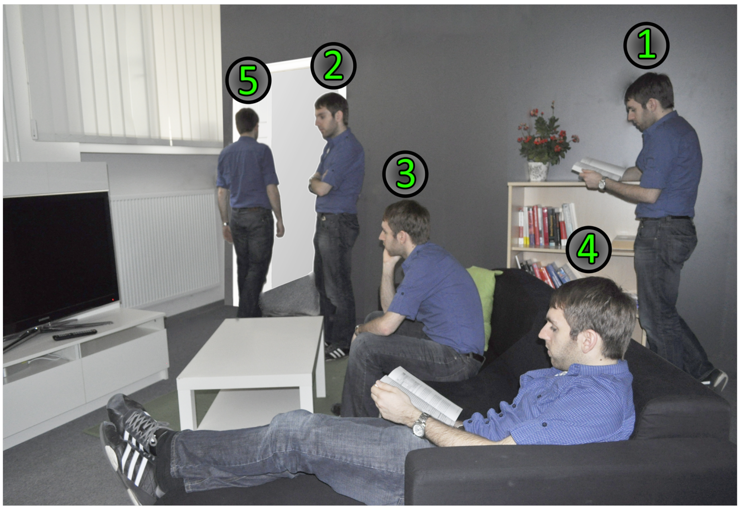


Figure 3.4: Fine-grained levels of engagement in TV watching activity include (1)Passive, (2)Inclined, (3)Involved, (4)Oriented and (5)Away.

Inclined is a level in which people stand in front of the TV, and their attention is also towards the TV screen. Based on our observations, in this level, it is very likely that people start zapping through TV channels in order to find a program.

Involved is when people sit in front of TV and engage in the content and their attention is mainly oriented to the TV.

Oriented is a level in which people are already involved in watching however, their attention is partially influenced by other activities that are the typical while watching (e.g., reading a newspaper) or by the presence of other viewers (e.g., a conversation).

Away is when people leave the TV room.

These levels of engagement are depicted in figure 3.4.

3.2.2 Design Requirements

Based on findings of the preliminary user study and results of literature analysis reported in 3.1, we derive four essential requirements as the rationale for our design.

R1. Support body-based coarse- and fine-grained TV interactions

Both the state-of-the-art examination and our study findings demonstrate that spatial knowledge of TV user's body, such as orientation, posture, and movement, can be exploited to design more effective and delightful interaction concepts. Since the TV interaction is still characterized by several deficiencies, such as requiring a mediator device, taking advantages of the body and its parts for coarse- and fine-grained TV interactions open ups opportunities to enhancing TV watching experience. Considering the human body as an interface and the hand as an interactive surface can potentially aid expanding the design space for implicit and explicit TV interactions, respectively.

R2. Support eyes-free TV input

We observed that since the TV screen takes full attention of viewers, alternating the users' gaze between the main screen, remote control, or probably touch-based secondary screen [Cesar 2009] can be interruptive and become cumbersome. For example, users need to switch their attention between the TV screen and button-based remote control [Hawkins 2005] or touch-based devices. As we have seen in the related work, several wearable and mobile interfaces used the surface of the hand and arm as input system. However, to the best of our knowledge, there exists no work that appropriates the hand as an eyes-free interactive surface leveraging the human proprioception sense.

R3. Support easy (re)engagement into watching activity

Our observation of people TV watching habits at home showed that when people are in front of the TV, they do not just watch the TV. We identified that the watching activity has various levels of engagement. This begins with the presence of a viewer in a living room in front of the TV until she gets fully engaged with the content. Although fully engaged with the program, viewers may change their posture or position for different purposes. They do not only attentively watch TV in a sit-down manner, but also, they may lie down to relax, lean forward to concentrate, or even temporarily leave the room to do something quickly. Moreover, there are environment-related interruptions, such as a phone call or starting a conversation with others, which may distract the viewers from watching and eventually, diminish the user experience.

Recalling the prior systems discussed in 4.1, none of them have supported easy (re)engagement into the watching activity. Later in this chapter, we propose that (re)engagement in TV content can be potentially supported through understanding and continuously detecting user spatial and postural movements.

R4. Support efficient overview and awareness

Requirements		Supported by prior work?	Contributions of CouchTV and PalmRC
R1	Support for body-based coarse- and fine-grained TV interactions	◐	CouchTV provides awareness and facilitates engaging in the watching activity. PalmRC supports navigating and selections of items in the TV interface.
R2	Support for eyes-free TV input	◐	Both interfaces do not require any attention while interacting with the TV.
R3	Support for easy (re)engagement into watching activity	○	CouchTV response to viewers (re)engagement into watching activity by supporting information about missing scenes.
R4	Support for efficient overview and awareness	◐	Through awareness techniques, CouchTV provides various levels of awareness about the program.

Table 3.2: Overview of design requirements. ○and ◐show if the state of the art have covered the requirements to some degree respectively.

Our observations and participant's feedback revealed that providing overview and awareness is important for a rich TV watching experience. For example, when returning back to the TV room, displaying "program time-line" or "friends' activities" may help to know whether it is still worthy to join watching the program together with other viewers. This can be achieved by displaying the overview and awareness of the program in the right way and at the right time leveraging user identity, presence, and particularly body posture. Otherwise it may become bothersome and eventually diminish the user experience.

Table 4.3 illustrates a summery of the requirements stated above. In addition, we show that weather they are addressed in the state-of-the-art and also how we aim to support them in the contribution of this chapter. The following section, presents the interfaces that we contribute to meet the above stated requirements.

3.3 CouchTV: Body-Based Interaction with TVs

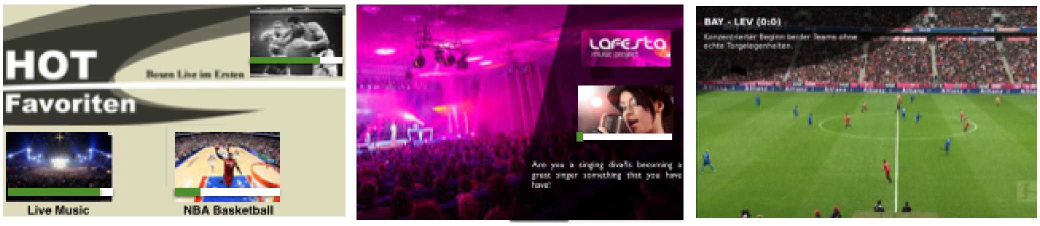
With regard to the design requirements established in 3.2.2, we present CouchTV, a body-based TV interface that infers the user's postural and spatial information at home and provides appropriate feedback. This proposes a set of body-based interaction concepts providing awareness and supporting implicit interaction for natural TV control. We present each of the interaction concepts in the next section and refer to the respective requirement that is supported. At the end, an early user feedback was conducted in which we investigated the CouchTV usage and its watching experience attractiveness. The results are reported in Section 3.3.3.

3.3.1 Underlying Interaction Concepts

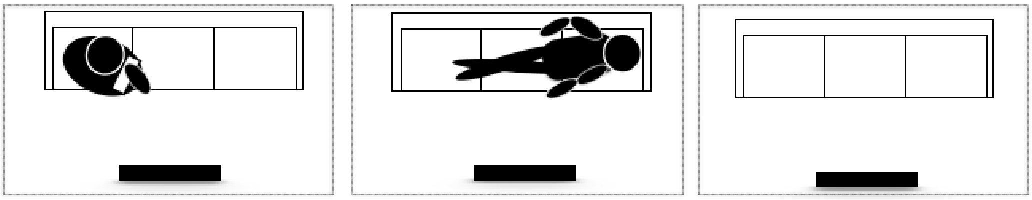
In this section, we describe two interaction concept –coined as awareness-supporting and implicit interaction – and the respective interaction techniques summarized in table 3.3.

3.3.1.1 Awareness-Supporting Concept

In response to the user's different level of engagement in the watching activity, CouchTV activates awareness that provides information about the TV content (R4). This relies on user's body movement and is aimed to be triggered when the user is sufficiently interested and find it necessary (i.e., inclined or involved levels of engagement 3.2.1.2). CoachTV supports displaying three types of awareness:



(a) Welcome awareness, Notification awareness and EPG awareness



(b) User's postural position while implicit control, suggest, and pause & record techniques

Figure 3.5: Awareness-supporting Interaction Techniques.

Welcome awareness visualizes information that is helpful to select what to watch, like current favorite programs based on user's profile, or which programs online friends are watching. This appears on TV as a person situates in the inclined level and facilitates seeking interesting programs to decide what to watch (R3). If the person leaves, the TV becomes inactive.

Notification awareness is designed to provide supplementary information about missing contents after viewers are distracted by an interruption or are away for a while. For instance, detail about who and when someone scored a goal in a football match appears on the corner of the display as soon as viewers' attention is redirected to the TV. This behavior is inspired by our observation in the preliminary study: viewers had difficulties reengaging in the program and distracted other coviewers to ask about missing scenes.

EPG awareness occurs once a program is finished and another one is started while viewers are interrupted. This supports high-level information about new programs, such as title, short description, actor's name, and timeline. This type of awareness helps a user to decide on continuing watching the newly started program (R3). This also allows viewers to benefit from the displayed information instead of seeking information in the Internet or EPG.

These awareness levels are depicted in figure 3.5.

3.3.1.2 Implicit Interaction Concept

Although the user's spatial and postural position are not primarily aimed to mediate interaction with TV, in this section, we show how this can be leveraged to operate TVs (cf. figure 3.5). Our concept aims to support implicit interactions with TV systems, namely implicit control, implicit suggest and implicit pause and record.

Implicit control controls the TV based on the focus of the user's attention while he is holding or facing toward any object or screen in the living-room (i.e., volume down while receiving a call).

Implicit suggest provides suggestion for implicit interaction to see if the system had a correct understanding of what the viewer is up to in order to minimize unwanted social interactions (i.e., activating the sleep mode while viewer is lying on the coach and not facing the TV screen anymore).

Implicit pause & record occurs when a viewer leave the room and the system implicitly reacts to this situation by pausing the time-shifted video and recording the live ones (i.e., record a part of a soccer match while viewer is out to bring a bowl of popcorn).

We demonstrate our envisioned interaction techniques via a walk-through scenario considering levels of engagements discussed in 3.2.1.2 and then present the interaction techniques.

Walk-through Scenario

Initially, the TV is inactive, and no one is in the TV room. Paul enters the room (passive) (cf. figure 3.6). The TV activates its screen and displays the Paul's favorite programs that are currently running on the TV and also provides an overview of Paul's online friends who are watching this program. Paul approaches and stands in front of the TV while looking at its screen (inclined). The TV displays more details about the program such as a timeline, which shows when it is started. Paul can sit on the couch, which causes the TV to play his favorite program in full screen, while directly zapping through other programs through the remote control or not paying more attention to the TV. Finally, he decides to sit and start watching his favorite boxing match with remote friends (involved). Later, he started watching a soccer match, but in the meantime, his wife asked him to come to the kitchen (away).

When Fred reenters the room while his attention is back to the TV, a notification appears on the corner of the TV screen about the goal that was scored while he was

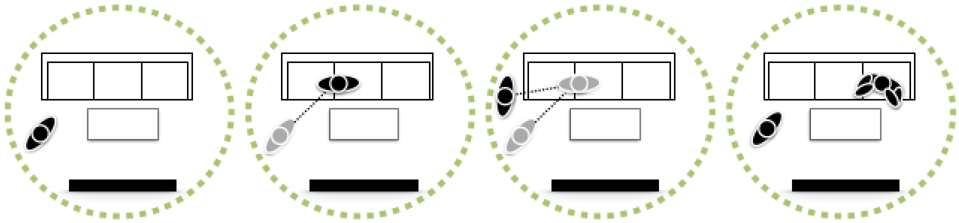


Figure 3.6: The user's position and postures described in the walkthrough scenario

away (involved). The TV also notifies that it records the missing scene for Paul. Later, while watching movie, Paul's wife enters the room to grab her book from the bookshelves. She looks at the TV display to see what Paul is watching (inclined). So the TV displays information about the programs that he is watching such as its title, short description and actor's name as it is shown in figure 3.5. It's noticeable that this information also appears if Paul reenters the TV room and the program changes during his absence. As she finds the program not interesting, she leaves the room (away). At this time, Paul receives a phone call (oriented). As he picks up his phone, the TV decreases the volume till the call is finished.

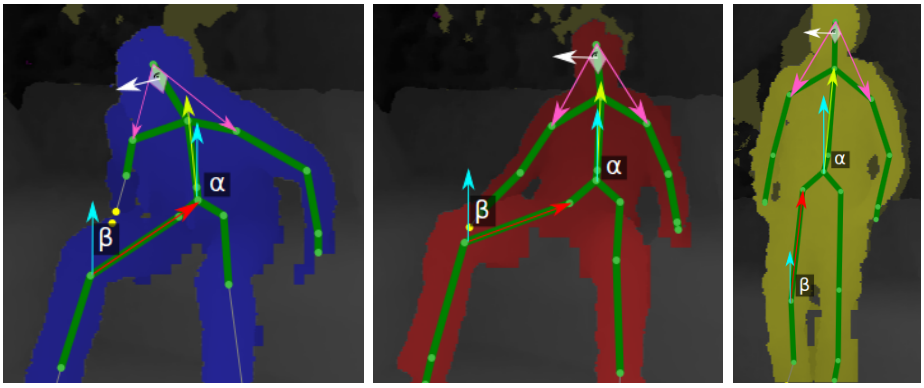
Interactions with CouchTV that are proposed based on the user's spatial body information, such as Paul's in the previous scenario, do not necessarily form an exhaustive set of implicit techniques. They provide solution to typical problems caused in different levels of engagement while watching. When a person is in the inclined level of involvement, the TV automatically gets activated. Once the person is involved in watching activity, the TV reacts based on her spatial situations. By receiving a phone call, the TV adjusts the volume. If the person lies down on the couch, the TV asks the person to activate sleep mode. A similar approach is taken if an involved person goes away. The TV pauses the programs. If the person stays relatively long away, the system inactivates TV.

In collaborative situation, TV records videos if a person leaves the TV environment to not interrupt others from TV watching. However, it supports the away person to the environment showing, for instance the recorded video in a picture-in-picture feature on the main screen or full screen on his secondary device upon his return. Through this way, the TV can utilize viewers' spatial situation in performing implicit interactions as a complementary input modality for the user's explicit TV commands.

Table 3.3 summarizes the interaction techniques proposed by CouchTV interface to support easy (re)engagement in watching activity and body-based TV interaction.

Awareness-supporting Concept		
Name	Purpose	Description
Welcome awareness	Encourages the user to get in touch through their televisions	Reacts based on the user's spatial position and displays related information (i.e., current favorite programs based on the user's profile, or which programs online friends are watching).
Notification awareness	Supports easy re-engagement into watching activity	Reacts after viewers are distracted by an interruption and displays related information (i.e., who scored a goal in a soccer match or who was the killer in a crime movie)
EPG awareness	Help as user to decide on continuing watching the newly started program	Reacts once a program is finished and another one is started while viewers are interrupted and displays related information (i.e., title, short description, actor's name, and timeline)
Implicit Interaction Concept		
Name	Purpose	Description
Implicit control	Enhances user experience	adjusts volume activating the TV set
Implicit suggest	Supports effort-less interaction	Activates TV sleep mode, inactivates the TV set
Implicit pause and record	Supports easy re-engagement to the TV content	Pauses time-shifted video broadcast, Records live broadcast

Table 3.3: Summary of the interaction concepts and the respective interaction techniques presented in the CouchTV prototype



(a) Leaning Forward (b) Leaning Back (c) Standing

Figure 3.7: CouchTV is able to recognize different user postures using the vectors between skeleton joints and their alignments in the 3D space.

3.3.2 Implementation

We developed our proof-of-concept CouchTV prototype using a Microsoft Kinect depth camera. It does not require any instrumentation on the viewers' body. We mount the depth camera on top of TV screen. The built-in depth sensor detects user in a distance between 1.2m and 3.5m to the TV display.

Postures are recognized in a multistep process based on vectors between skeleton joints and their alignments in the 3D space. We calculate three vectors, one from user's upper body (hip to neck), one from user's lower body (knee to hip), and the other as a normal vector straight up from the floor to the ceiling depicted as yellow, red, and blue vectors in figure 3.7, respectively.

To determine different postures of the user, we compare the angle (alpha and beta) between these two vectors with the normal vector. In order to calculate the user's attention toward the TV, we observe the normal vector of the plane span by the joint positions of the user's spine and the two shoulders (cf. figure 3.7). If this hits the TV position, then the attention of the user is oriented toward the TV screen. Using this tracking data, CouchTV reacts to the user's different spatial situations (explained in3.3.1.2) and displays appropriate awareness levels on TV screen or proposes implicit interactions (cf. figure 3.5).

3.3.3 Initial User Feedback

We conducted a study to understand if CouchTV can enhance TV watching experiences. The study took place in a lab, furnished like a living room to simulate real-life situations.

3.3.3.1 Study Design and Methodology

We recruited 12 groups of two friends (10 male, 14 female, avg. 25 years, SD.: 7.2 years). Fifteen participants were students, six of them are faculty members or university staff, and the rest did not list their affiliation. All of them but two were regular TV viewers. They were introduced to CouchTV upfront.

The study contained two watching conditions: (1) with CouchTV system reactions and (2) without any system support (normal TV system). The study took about three hours for each group. The order of conditions was counter-balanced across groups. We picked two shows from five different genres (sitcom, cooking show, drama, sport, and news) to study the effect of genre on our concept.

Since the result of the pilot study showed that interruptions from watching activity caused people to transit between different levels of engagement in watching activity, one of the study observers of this paper participated in the study to simulate the real-life interruptions during the study. He did not attend in watching activity but externally imposed participants some interruptions, such as asking questions, making a phone call or offering some drinks, and snacks, in both conditions of the study.

The goal of these interruptions was to cause participants either to leave the room for a short while or to lose the focus from TV content. Moreover, we let participants answer their phone calls as well as leave the TV room to go to the restroom or to smoke cigarettes during the study. These interruptions and different spatial situations of participants, such as their body postures, head orientation, etc., in front of the TV enabled us to showcase our system's interaction techniques.

3.3.3.2 Data Gathering and Analysis

We used interaction logs, video and audio recordings, and the AttrakDiff test [Hassenzahl 2003]. Moreover, six focus group sessions were conducted after each study with two groups of participants. The focus groups and videos were transcribed and analyzed using an open coding approach [Strauss 1998]. The AttrakDiff test contained a list of extreme antonym attributes, characterizing a user's perception of the watching experience with both normal TV and our proposed system. The three following dimensions of user experience could be evaluated by this questionnaire:

- Attractiveness (ATT): Describes a global value of the product based on the quality perception.

- Pragmatic Quality (PQ): Describes the usability of the system experience and indicates how successfully users are in achieving their goals while experiencing using the system.
- Hedonic Quality (HQ):

  Identity (HQ-I): Indicates to what extent the experience allows the user to identify with it.

  Stimulation (HQ-S): indicates to what extent the experience can support those needs in terms of novel, interesting, and stimulating functions, contents, and interaction- and presentation-styles.

This test was used to evaluate all these three dimensions and the overall user experience while watching TV using CouchTV system. It was noticeable that this test measures the hedonic and pragmatic qualities independently; therefore, they could equally affect the rating of attractiveness [Hassenzahl 2003].

Our participants completed two AttrakDiff questionnaires immediately after each session to avoid difficulties in recall. In the following, the results of the overall user experience evaluation and preferences are presented.

3.3.3.3 Results and Discussion

Here, we reported and discussed our research findings related to the general reactions of users to our prototype and the overall TV watching experience while using both normal TV and CouchTV.

AttrakDiff

We calculated the mean score of all answers for each quality scale in AttrakDiff test (each scale has seven questions). The overall feedback to our concept was quite positive. Figure 3.8 illustrates all the quality scales have means above zero in both sessions with or without simulation.

The results showed that our concept can significantly enhance hedonic quality by 95% certainty where there were no incidental fluctuations. Similarly, almost all participants' quotes confirmed that they enjoyed experiencing our concept, and this is a way for future TV, but only P22 said *"It's a little scary as you feel someone is in the room and knows what you are doing."*

For questionnaires, a 2D graphic is generally generated, which places the CouchTV and typical TV systems in terms of PQ (i.e., usability of the system) and HQ (i.e., novelty, interest, etc.). The typical TV system was qualified as task-oriented. It

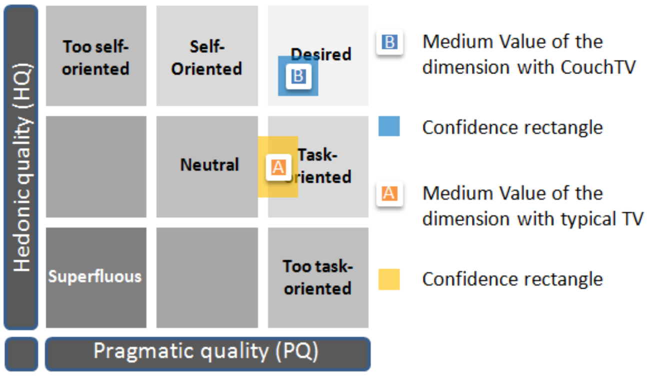


Figure 3.8: Results of AttrakDiff comparing watching experience using typical television vs. CouchTV

means that the typical TV was usable but it was rated average in terms of HQ. The confidence rectangle was particularly large in the hedonic dimension, which means that participants' opinions were mixed.

In terms of attractiveness, the overall impression was that normal TV is moderately attractive. According to our participants, watching typical TV was "manageable" and "simple", but also very "ordinary" and "isolating." The watching experience using the CouchTV system was qualified as desired. It means that CouchTV optimally assisted and motivated the users.

The confidence rectangle was small for both HQ and PQ. In terms of attractiveness, the overall impression was that this system is highly attractive. Participants found it very "novel", "pleasant" and "presentable," but the bad point was that it was not challenging for the users.

General Observations

Here, we report the issues identifies while study observation and particular things that participants appreciate while using CouchTV. The results revealed that CouchTV:

1. *Fitted as an ambient display at home*: Participants appreciated activating TV screen and receiving awareness about their favorite programs and online friends. P6 commented *"that (Initial awareness) is great because I enter the TV room several times a day for various reasons. So I can be aware if any of my favorite program is showing on TV and don't miss them"* and P19 said *"If I had this feature on my own TV, I explicitly pass the TV room to just know which program is currently going on"* and P1 added, *"I hope I could also check my Facebook and news in addition*

to TV program in the morning just by looking at TV screen". There already exists technology, which supports this to some extent. However, it does not consider TV interaction based on viewers' spatial information.

2. Supported effortless information gathering while watching: The majority of participants indicated that the content awareness which appears on TV screen after interruptions, enabled them to easily and immediately reengage in the TV program without looking up themselves. P2 said *"The awareness about new program that began during my absence is really useful as it helped me to decide to watch it or zap to another program.*"

3. Enriched interactivity with TV systems: According to our participants, CouchTV prototype enriches the TV watching experience in terms of interactivity. Participants commented that they frequently came across situations in which they desired using no 3rd party mediator device to operate TV as it was lost (e.g., P1 stated, *"I'd like recording feature while I was watching the football match when I had to open the door for my friend when he was knocking it"*) or it was misplaced (e.g., P24 stated, *"I was lying on sofa and my phone was ringing. I wanted to ask my friend to reduce the volume but TV just did it when I picked my phone"*).

4. Required to be synced with on-screen content: Our primary research question was whether content awareness during program is seen as a useful addition to the TV viewing experience. Participants confirmed that with their comments during the focus group sessions. They clearly indicated that their experience is positive only when the relevant information is in sync with the TV show and appears as soon as the user looks back to the TV screen to reengage with the program content.

P23 said, *"The awareness was definitely helpful as it was immediately there, right when I came back to the room and was curious about what happened during my absence!"*. P8 put it more emphatically saying *"The system should make sure that the awareness is in sync with the content. I do not want to receive any info about an event which did not happen yet or is happening currently in the program.*"

5. Supported social and content connectedness: the participants emphasized that they enjoyed the watching experience, while CouchTV provided awareness for whomever else reentered the room, and thus, they were not distracted from watching. P18 commented, *" I definitely liked it because my partner read what has happened when he was not at room and did not ask me when I was focusing on a movie* *dialogue*." However, CouchTV also induced a limpness between coviewers: the more they received content awareness through CouchTV, the more they were disconnected from coviewers.

Nevertheless, the actual disconnection is not to be regarded as pure negative, since intentional disconnection, such as P18 statement, can enhance the overall experience. Additionally, several participants mentioned that they frequently came across situations in which they entered the TV room and wanted to look up about and probably join the TV shows, which another collocated person was watching. However, they failed to do so because asking them could be too distracting for the viewers or the viewers would response only to one or two questions with incomplete answers as they were deeply involved in watching activity.

6. Highlighted the role of genre on awareness consumption during program: The participants mentioned that content awareness should be adapted to genre with different number of plots, video length and content quality. P2 said, *"The importance of awareness can be really dependent on what I'm watching. For example, if I have to leave the TV room for two minutes, I may miss much more information while watching breaking news comparing to watching movies*." Some participants also noted that they want to receive awareness for a rich-ingquality video which asks for viewers' continues attention and they do not care about what they miss about poor programs which are not that interest.

Table 3.4 summarizes the main findings of the CouchTV study.

In summary, we have presented our approach to appropriate the human body's spatial and postural information to enhance interaction with TV systems. We have described a body-based sensing television system that response to user spatial situation by providing awareness and proposing implicit interactions. Initial user feedback have shown that our system is appreciated and can aid to ease (re)engagement into watching activity and coarse-grained implicit TV interactions. In the following section, we move to focus on investigating fine-grained body-based interactions that support omnipresent, deviceless, and eyes-free TV input modality.

3.4 PalmRC: Palm-based Interaction with TVs

As it was discussed in section 3.2.2, TV interaction is still characterized by several drawbacks. TV input is predominantly supported through remote controls. Common examples are button-based conventional remotes or touch-based interfaces on smartphones. Thus, users are always required to utilize a particular mediator device

Evaluation Results
– CouchTV resembled an ambient display at home.
– Obtaining information about TV contents while watching was found to be more effective using CouchTV system.
– The proposed interaction techniques were found to be intuitive, practical, and enjoyable.
– The proposed awareness-supporting concepts were required to be synced with user spatial situations to enhance watching experience.
– CouchTV found a limpness between coviewers: receiving more content awareness through CouchTV and becoming more disconnected from coviewers.
– The usefullness of awareness-supporting techniques was found to be increased while watching genre with better content quality.

Table 3.4: Summary of the main findings from the CouchTV evaluation as an body-aware TV user interface

to interact with the TV. While admittedly a well-established interaction paradigm, it has various drawbacks. On the one hand, the device itself can be out of reach or misplaced or even lost [Bernhaupt 2008].

In addition, users typically have to deal with several remote controls for different home entertainment devices, each with an excessive number of functions assigned to various physical buttons. This makes remote controls even more complicated and confusing than before [Bernhaupt 2008]. On the other hand, touch-based interfaces on mobile devices require a lot of attention and users have to constantly switch their attention between the device and the content on the TV [Hawkins 2005]. This increases a user's effort for controlling the TV and therefore diminishes the user experience while watching.

In response to the key requirements R1 and R2, we propose PalmRC, a novel imaginary body-based interface to operate TV systems. PalmRC transforms the palmar of the nondominant hand into an interactive input surface. Users can then operate the TV through touching the palmar with the other hand's index finger. PalmRC builds on the sense of proprioception [Li 2010]: humans are unconsciously aware of the relative position and orientation of their own hands. Therefore, the palm can be appropriated for eyes-free TV interactions. In consequence, PalmRC neither demands a user's visual attention nor requires an additional mediator device.

We explore the concept of PalmRC in a series of user studies. We begin with describing the first study, being exploratory in nature. We aimed at gaining insights into the conceptual space of palm-based remote controls. In the following section, we particularly investigated different interaction styles and elicit implications on how to design such remote controls. Based on the results of this study, in Section 3.4.2, we present a controlled experiment that investigate the human capability of touching owns palm without paying any visual attention. More precisely, we aim to quantitatively answer the following questions:

- How precisely can users touch their palm's salient regions (landmarks) without looking at them?
- How effectively can they select the target element of transferred on-screen user interface elements on their palm by pointing to the corresponding region on its surface without any visual attention?

We then conducted a third study which is explained in section 3.4.5. There, we compare PalmRC to conventional button-based and touch-based remote controls. In this study, we particularly focus on identifying respective advantages and disadvantages of each input modality.

The results of all three studies provide deep and broad insights into the conceptual design space and clarify concrete design questions like precision, effectiveness, and user experience pertaining to the concept of PalmRC. Later, as a proof of concept, we designed and implemented a functional prototype using depth sensing technology. Our prototype supports common tasks such as zapping through channels, menu navigation or social interaction between remote viewers.

3.4.1 Study1: Exploratory Experiment

We initially explain the exploratory study to empirically ground the requirements for designing an eyes-free, palm-based TV remote control. We were particularly interested to see how users would interact with their hand to perform a set of common interactions with TVs, while preserving their attention to the TV screen.

3.4.1.1 Study design and methodology

The study had a brainstorming character in which participants were asked to discuss high-level aspects of using the palm as a remote control. Initially, we asked participants to particularly elaborate on:

Figure 3.9: Example of user interfaces of Samsung TV used in the exploratory study.

- How would participants hold their hand and which side and parts of their hand would be suitable for interacting with the TV?
- How would they transfer the remote control functions on their hand?
- How would they interact with on-screen UI elements while mimicking their proposed interactions on their hand surface?

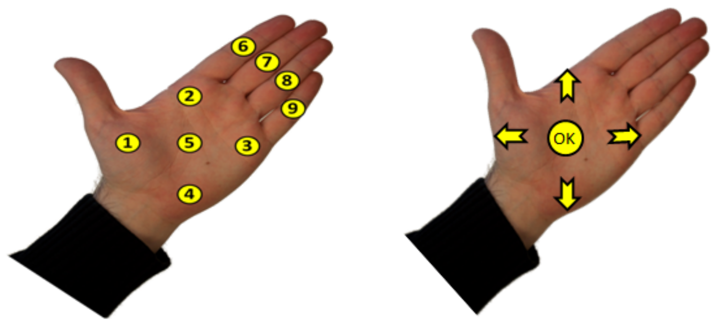
To foster the discussion, we utilized and displayed some typical user interfaces of a Samsung Smart Internet TV on its screen and asked participants to show how they would interact with these elements using their hand. The user interface screens can be classified into three vertical, horizontal and whole screen grid-based menus (cf. figure.3.9).

We recruited ten volunteer participants (three female). They were between 22-42 years old. All participants spent 2-3 hours in average per day watching TV. Each single-user session lasted about 1 hour. As data gathering methodologies, we videotaped the sessions and asked participants to think aloud. We then selected salient quotes and analyzed both quotes and videos using an iterative open, axial and selective coding approach [Strauss 1998]. For inter-coder independence, two coders coded the data separately.

3.4.1.2 Results and Discussion

Using the Palm surface as a TV remote control

Generally, participants appreciated the idea of being able to use the palm surface for operating the TV. Unlike the one-hand usage of typical remote controls, all participants used their palmar (inner side) of the nondominant hand as an input surface and interacted with the other hand's index finger similar to [Gustafson 2011]. They



(a) (b)

Figure 3.10: (a) The easily touchable landmarks on the hand. (b) Participants suggested linking the directional keys to the landmarks of the palm while holding the hand diagonally.

said interacting with the palmar is not only more intuitive, but it also offers several salient regions (landmarks) to easily interact without any visual demand. P3 said "I am able to properly touch any of my fingers as easy as moving them." and P8 added "I can touch four curved areas (convex) on my palm surface even in the darkness". Participants revealed nine landmarks on the palm surface, which they believed to be easily touchable without any visual demand based on the proprioception sense [Li 2010] (cf. figure3.10).

Mapping basic remote controls functionalities

Participants mentioned that they would only map frequently used functions to their palm such as navigation, selection, digits for direct switching between channels, volume adjustment, or play and pause. In addition, they offered to properly map these functionalities to the location of landmarks of the palm, since they can be easily hit without any visual attention. For example, participants stated that the mapping of directional keys could exactly match the four convex and one concave landmarks of the palm (cf. figure3.10).

In contrast, recalling and transferring digits (typical mapping of 3x3 buttons of digits from 1 to 9) to the palm was found to be very complex. P5 said "Digits may have a conventional mapping but still they lack having a natural mapping and I would prefer to draw digits on my palm to change the channels". P7 added: "Even if I could recall each digit position, I would not know where to map it on the palm surface as no landmarks afford their mapping". Participants also commented, since

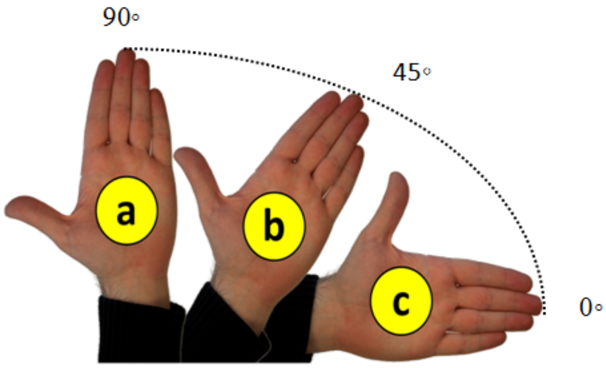


Figure 3.11: (a) Portrait: pointing toward TV. (b) Diagonal: 45 to user's body. (c) Landscape: Parallel to body.

no digital information is projected on the palm surface, the simplicity of the design of a palm-based remote control is crucial.

Interacting with on-screen UI content

Participants not only suggested 2D-touch gestures (e.g. swipe, scroll, and draw) on the palm, but they also proposed mapping UI elements displayed on the TV screen to the palm's surface. They then imagined triggering the target elements by pointing (tap, click) to the corresponding location on the palm surface. For this purpose, participants used three different hand orientations including diagonal, landscape and portrait (cf.figure3.11).

The diagonal orientation was stated as the most comfortable form of holding the hand as an interactive surface. The interactions requiring participants to map remote control functions to their palm (such as directional keys), as well as 2D-touch gestures, were mainly performed in diagonal orientation.

2D-touch gesture interaction on the palm surface

Although a palm is not a flat planar surface, participants considered it as a concrete surface and proposed using 2D touch gestures on it while holding it in diagonal mode. This interaction technique was typically proposed for either efficiently browsing menus with a plethora of options, or mimicking digits on the palm surface for channel navigation, or even nonverbal communication between remote viewers; as P3 stated, "I could for instance draw a smiley on my palm surface and send it to my online friends who are watching the same program".

Pointing on the palm surface
Participants suggested to transfer one-dimensional grid-based UI elements (e.g. list of applications or media player controls with three buttons including backward, pause and forward) onto the palm surface. While looking at the TV, participants first mapped the whole screen of the UI to the nondominant hand surface and then selected/triggered UI elements by pointing to the corresponding location on the hand surface using the index finger of their dominant hand. Participants transferred the grid-based vertical and horizontal UI screens to their palm while holding it in portrait or landscape orientations respectively.

Participant's comments highlighted the fact that the design of TV UIs elements based on the location of the palm landmarks may improve the mapping. P4 stated, "If a menu could have four options, I could easily touch my middle finger to select the second option". Discussion with participants revealed that hand-tailored TV UIs may decrease the cognitive effort of mapping these elements to the surface of palm and eventually results in more secured feeling of hitting appropriate location on the palm while looking at the TV.

3.4.1.3 Summary

The results of this study elicit implications for designing a palm-based remote control, which preserves a user's attention to the TV screen during interaction. Table3.5 summarizes the findings from our exploratory study presented above. The results lead to two unexplored questions related to R1 andR2 discussed in Section 3.2.2:

- How precisely can users touch their palm's salient regions (landmarks) without looking at them?
- How effectively can they select the target element of transferred on-screen user interface elements on their palm by pointing to the corresponding region on the palm surface without any visual attention?

3.4.2 Study2: Controlled Experiment

The aforementioned questions in the previous section were formulated as hypotheses of another study. We have verified these hypotheses in a controlled experiment. The two questions map to the following two hypotheses:

- H1: People can touch their palm landmarks precisely without looking at them (0.90 confidence level).

General Findings
– Users preferred to transfer typical remote control functionalities to the surface of their palm.
– Nine distinct landmarks were found on the palm surface that can be easily touched without visual attention.
–The landmarks can be linked to the common TV functions of a remote control (e.g., directional keys).
– Users preferred 2D touch gestures for efficient browsing of lists with so many options.
– Users utilized the palm surface as a canvas to draw short symbols such as digits or emoticons.
– Considering the orientations of the hand, the visualized interface elements on the TV screen can be tailored to the hand orientation.
– TV users can switch between different menus based on the orientation of the hand.

Table 3.5: Summary of the main findings from PalmRC exploratory study.

- H2: When mapping on-screen UI elements to palm,

  H2.1: the effectiveness will decrease, the denser the UI elements are placed.

  H2.2: the effectiveness is independent of the UI elements' alignment; i.e., whether they are horizontally or vertically aligned.

Effectiveness here means whether a participant successfully touches mapped UI elements on her palm.

3.4.2.1 Experiment setup

We conducted the experiment supported by an optical tracking system to minimize any noise (cf. figure 3.12). This type of optical systems include two or more cameras that provide an overlapping projection field and three or more especial markers attached to a target (i.e., object or user) to identify the 3D position, motion and orientation. The marker-based tracking system may use different type of markers for various purposes. In this experiment, we used retro-reflective markers that can reflect the light back to the infrared cameras (IR), which high accurately calculates the potential marker positions.

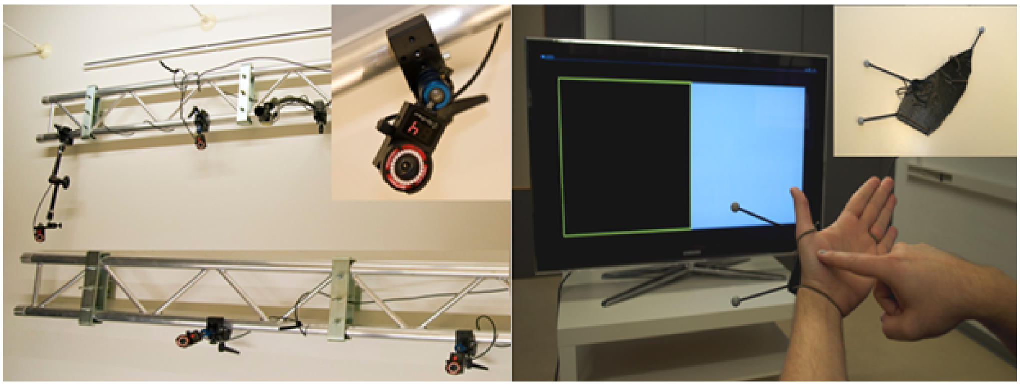


Figure 3.12: Left:OptiTrack system. Right: The paper carton apparatus used in the controlled experiment.

In addition, we have designed a trackable paper carton apparatus as the target, which the participants wore on the back of their nondominant hand (cf. figure 3.12). Hence, we have attached three retroreflective markers as antennas to the paper carton. These markers are then tracked by the OptiTrack system with six IR-cameras and define a 3D plane that corresponds to the palm surface. This allowed us to reliably track the palm without covering the palm completely, (e.g., using a glove in real time). To allow for accurate touch input on the nondominant hand, we have augmented the index finger of the dominant hand with another marker. A touch then is calculated by projecting the marker position on the hand plane and measuring the distance. We recruited 15 participants (5f, 10m; 32 years of age in average, with near-to-perfect sight). The participants were introduced to the system upfront. Each single-user session lasted about 45 minutes.

3.4.2.2 Methodology

The experiment was subdivided into two parts according to our hypotheses. Each part was again subdivided into two tasks (cf. figure 3.13). The order of the presented targets within each task was completely counterbalanced. The system advanced to the next target after each touch, regardless of whether the participant had successfully touched the target. We chose a within-subject design. Participants were asked to not look at their hands and only concentrate on the interface shown on the TV screen. We only repeated the trials in which the experimenter determined that participant looked at her palm.

Part I: Precision

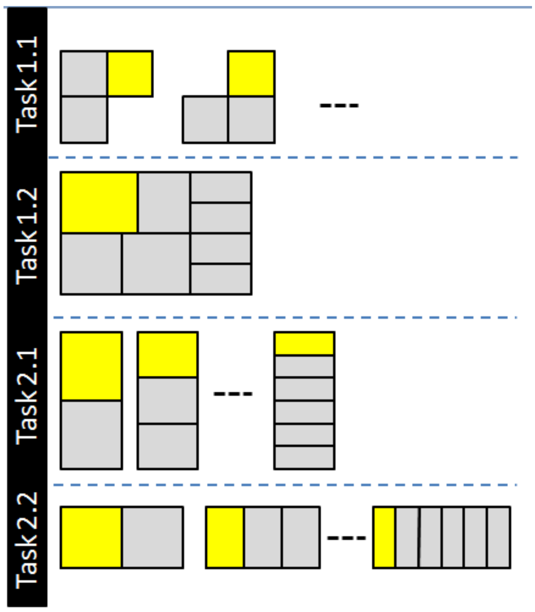


Figure 3.13: On-screen user interfaces of each task during the experiment.

In the first part, participants were asked to touch landmarks without visual attention. Independent variable was the landmark location. Dependent variable was the success rate of a user touching the landmark on her palm. Task 1 was comprised of two sub-tasks.

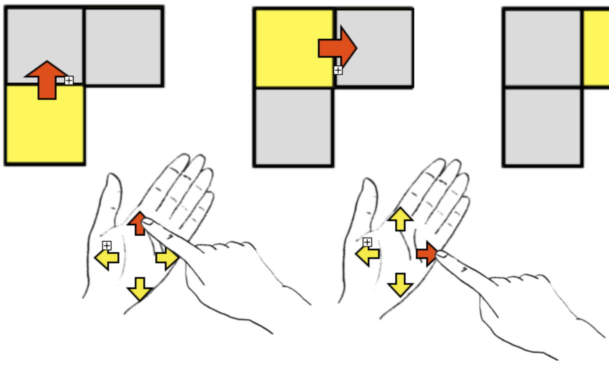
- Task 1.1 required participants to map directional keys to their palm (see figure 3.10), and navigate through a path of target items starting from the highlighted one (yellow box). For example, the first layout of task 1.1 in figure 3.13) required the participant to first touch left, then down. Participants had to touch 9 different landmarks.

- Task 1.2 required participants to map non-regular grids (see figure 3.14) to their palm surface and touch the highlighted position on their palm. Here, participants had to touch 8 different landmarks.

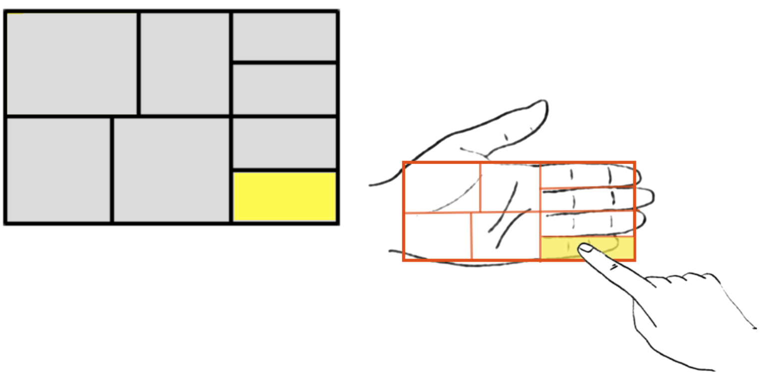
Part II: Effectiveness

In the second part, participants had to map and touch UI elements on their palm surface. Independent variable was the on-screen layout. Again, dependent variable was the success rate of a user touching the landmark corresponding to the UI element on her palm.

Task 2.1 required participants to map vertical 1D regular grids to their palm surface



(a) Task1.1



(b) Task1.2

Figure 3.14: On-screen user interfaces of each task during the experiment.

and touch the highlighted position on their palm. Each user had to touch 20 different targets.

Task 2.2 required participants to do the same with horizontal 1D regular grids, again for 20 different targets.

In order to determine boundaries for the number of targets in this task, we conducted a pilot study. We ask participants to target elements in various density levels starting from 2 adjacent targets in both horizontal and vertical orientations. We determined that participants were able to divide and eyes-freely touch the palm surface up to 6 locations at most. Therefore, the task started with 2 adjacent targets and stepwise became denser until 6 targets as depicted in figure 3.13.

Figure 3.15: Distribution of raw data of all participants by 90% confidence ellipses.

3.4.2.3 Results

Each target was repeated 3 times, leading to a total of 2565 data points over all 15 participants: 15 * 3 * [9 (T1.1) + 8 (T1.2) + 20 (T2.1) + 20 (T2.2)]. We discarded 21 trials as outliers, since they were farther than 3 times the standard deviation away from the centroid. We normalized all hand sizes with the average index finger (7.31cm).

Results I: Precision

Figure 3.15 shows the distribution of the raw data for tasks 1.1 and 1.2 by 90% confidence ellipses. This illustrates the spatial precision of the touches with respect to the centroid of each landmark. To analyze targeting, we measured one overall systematic error (offset). On average, the diameter necessary to encompass 90% of all touches is 28mm (SD= 0.85).

Results II: Effectiveness

The average effectiveness for each landmark is shown in figure 3.16. All of the palm landmarks were effectively touched with at least 94%. The finger landmarks were less effectively touched with as little as 53% for the pinky.

ANOVA tests revealed that the difference between palm and finger landmarks is statistically significant ($p<0.001$). Bonferroni post-hoc tests confirmed that this holds for all comparisons ($p<0.001$).

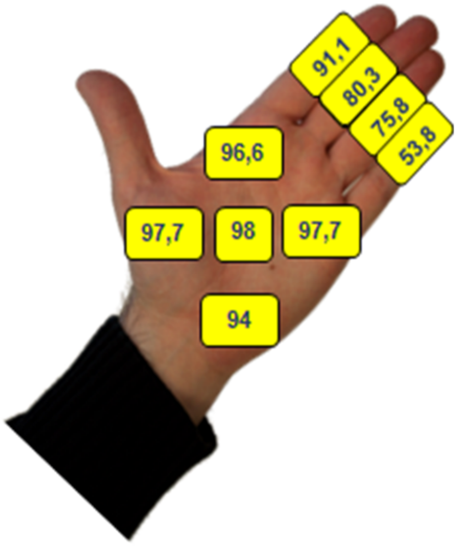


Figure 3.16: Average effectiveness percentage of targeting each landmark without visual demand.

The average effectiveness for the target elements is shown in Figure 3.17. The effectiveness decreased monotonically for more than 3 menu options. The average effectiveness is below 90% for more than 4 options and decreases below 50% for more than 5 options.

ANOVA with Bonferroni post-hoc tests revealed that these effects are statistically significant ($p<0.05$). The differences between horizontal and vertical alignments were not significant.

3.4.2.4 Summary and Discussion

Based on the results of the studies, we showed that touching the 5 landmarks on the palm surface without any visual demand is highly effective. Moreover, it is precise enough to operate interfaces with target sizes of 28mm in diameter (H1).

This implies that future palm-based TV interfaces should not map functions to regions with a smaller diameter. Moreover, this shows that users can effectively map common functions of traditional remote controls such as navigational keys to the landmarks of a palm and touch them to operate a TV.

Our results provide evidence that people can reliably and effectively (>90%) map 1D grid-layout menus with up to 4 options to their palm surface (H2.1), independent

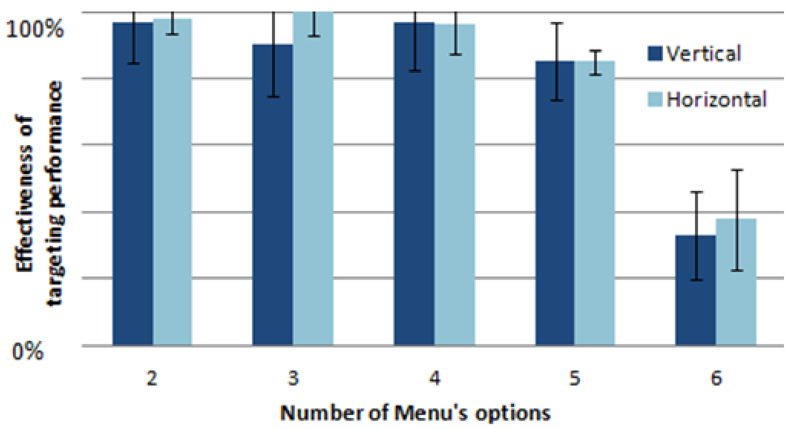


Figure 3.17: Average effectiveness percentage of targeting vertical and horizontal grids with different equal-sized options.

General Findings
– Touching the five landmarks on the palm surface without any visual demand was highly effective.
– It was precise enough to operate interfaces with target sizes of 28mm in diameter.
– People could reliably and effectively (>90%) map 1D grid-layout menus with up to four options to their palm surface.
– The high effectiveness in mapping and targeting equal-sized option of menus on the TV screen was independent of whether they were displayed horizontally or vertically aligned.

Table 3.6: Summary of the main findings from PalmRC controlled experiment study

of whether the menu is horizontally or vertically aligned (H2.2).

For future palm-based TV interfaces, we envision this to be leveraged as region-based shortcuts. While the participants were not as effective when touching their fingers compared to their palm landmarks, they effectively targeted their index finger. This indicates that also the index finger could be used as an effective input source. The findings of the controlled experiment is summarized in table 3.6.

In summary, the findings of the two studies presented in section 3.4.1 and 3.4.2 showed that users are able to use their palm to interact with TV without visual attention in two main ways: first, as a remote control with several functions (virtual buttons) that are linked to the landmarks and second, as an unique input surface

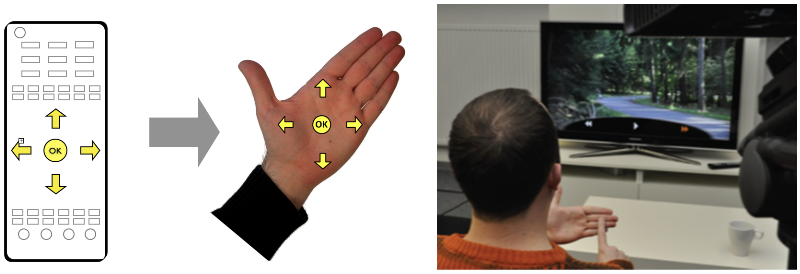


Figure 3.18: PalmRC: Mapping common functions of remote cotrols –e-g-, adjusting volume.

which the television user interface is mapped to the entire surface of the palm. Our findings showed that under certain circumstances (28mm button size and 4 target options) the palm-based remote control is viable. Thus, frequently used functions can be ready at palm, virtually any time without need for an additional mediator device.

Building upon these results, we developed a palm-based remote control called PalmRC along with two main interaction techniques: *linking functions to the palm's landmark and direct interacting with interface elements on TV screen*. In addition, we implemented several applications to show the usefulness of PalmRC for varying interaction scenarios with TV systems.

3.4.3 PalmRC User Interface

PalmRC allows users to operate the TV using empty hands while focusing their visual attention on the TV screen. The users interact by pointing and swiping on their nondominant hand and the system enables the surface of the palm to be capitalized as an input surface. The TV system receives touch positions and returns appropriate visual feedback on its screen. We developed interaction techniques to perform conventional TV interactions such as channel navigation in EPG, volume adjustment and direct interaction with menu options. PalmRC enables users to use their palm for various typical commands instead of retrieving a TV remote control.

3.4.3.1 Interaction techniques

PalmRC supports two main interaction techniques (modes) that make use of pointing and 2D-touch gestures on the palm surface:

Linking Functions to the Palm's Landmark

Figure 3.19: PalmRC: Chat and media-player application scenarios.

Based on the results of the first study, the diagonal orientation of the nondominant hand was found to be comfortable and resembles the style of holding a remote control in hand. Therefore, in this orientation, PalmRC links the common buttons of the remote control to the nine landmarks of the palm. Users can trigger buttons by touching the corresponding locations on their palm.

We implemented this mode for directional keys and a confirmation/menu button (as the most frequent used buttons). These are in turn linked to the landmarks of the palm, as revealed in our studies respectively. This technique also allows for a natural spatial mapping between the buttons and the landmarks. Users can zap through the TV channel by tapping on the corresponding landmarks, which are mapped to the up or down keys. The volume can be adjusted similarly by touching the locations of the right and left keys (cf. figure 3.18). To open up the channel list or menu options users can touch the center of their palm surface. Similar to touch-enabled devices, swiping upwards or downwards on the palm surface allows for a fast browsing of the channel list. Users can also directly switch to another channel by drawing its number on the palm surface.

Direct Interacting with Interface Elements

PalmRC directly maps the user interface screen to the entire palm or hand surface. So that users can touch the corresponding location of a target element on the palm. This interaction mode allows users to directly select a target on the TV screen.

We showcased this technique in a social interactive television interface, which incorporates common social features such as live chat (cf. figure 3.19(a)). Once users hold their hands in landscape orientation, the communication mode will be activated and they can directly select and interact with one of the options. We also

Figure 3.20: PalmRC: quiz show application scenario.

integrated this interaction technique in an application enabling remote viewers to answer questions of a quiz by pointing to the appropriate location of their palm (cf. figure 3.20). The technique provides quick and immediate interactions with the social TV interface.

As another application example, while watching a movie or a program, users can hold their hands in landscape orientation. Thus, the media player menu including three options as backward, pause/play and forward appears on the TV screen. Then users can map it to the palm and touch the corresponding location of the desired option (cf. figure 3.19(b)).

The current implementation of the PalmRC prototype support limited number of functions that are mapped to the hand's surface and its landmarks, resulting in a simple and unified user interface design. This is inline with the findings of prior studies showing that people want to reduce the overall number of remote controls and the number of keys on each[Pemberton 2003] . At the same time, the navigational and direct selection techniques offered in PalmRC can support and cover a wide range of common interactions with TVs [Bernhaupt 2008]. We, however, believe that the design space of PalmRC has a great potential to support advanced interactions. As an example, the number drawing feature on the palm can be extended in a way that users draw characters to enter text to the TV. Moreover, finger joints of the hand can be also leveraged as landmarks to map additional functions [Gustafson 2011].

Table 3.7 provides an overview of the interaction concepts and related interaction

Linking functions to the palm's landmark		
Name	Purpose	Description
Directional Keys and Confirmation/menu Button	Support natural spatial mapping between the buttons and the hand-landmarks	User can trigger buttons by touching the corresponding locations on their palm and holding the hand in diagonal mode
Swiping	Support fast browsing	Similar to touch-enabled devices, swiping upwards or downwards on the palm surface allows for a fast browsing of the channel list
Numbers	Support direct navigation	User can directly switch to another channel by drawing its number on the palm surface
Direct interacting with interface elements		
Name	Purpose	Description
Menu Mapping	Support direct and immediate selection of on-screen targets	User can touch the corre- sponding location of a target element on the palm
Meida Player Mapping	Support Quick video interaction	User can backward, pause/play and forward the video using media player menu on the TV screen

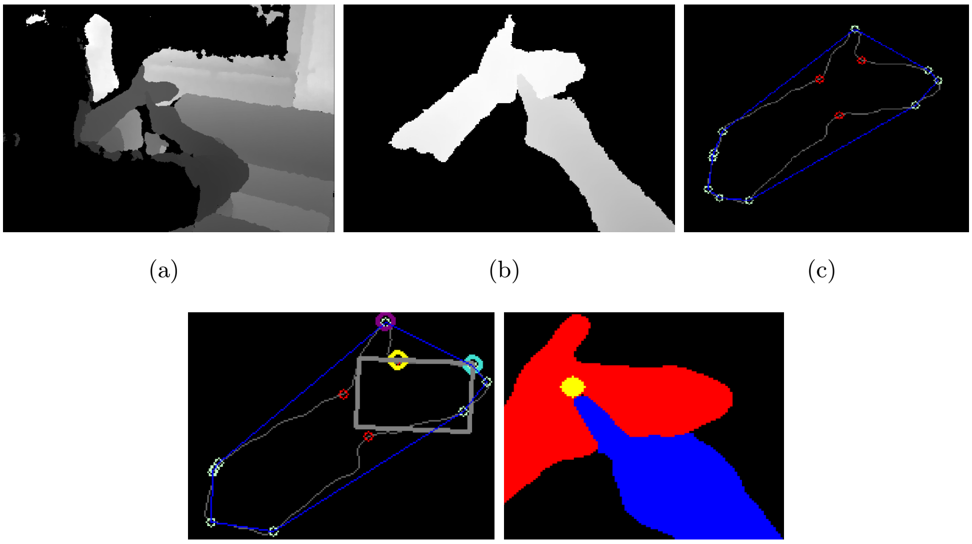
Table 3.7: Summary of the interaction concepts and the respective interaction techniques presented in the PalmRC prototype

techniques proposed by PalmRC interface.

3.4.4 Implementation

Although the OptiTrack motion capture system used in the controlled experiment and the comparative study enabled us to precisely track the palm and recognize the touch position, it is not practical for TV room settings. As discussed in Section 4.1, there have been other sensing approaches – such as gloves [Kuester 2005], Skinput [Harrison 2010], and depth cameras. We chose to use a Microsoft Kinect depth camera because it does not need any instrumentation on the hand of the viewer and also enables and supports recognizing touch and drag interactions [Gustafson 2011].

In PalmRC, we use the Kinect depth camera to track the nondominant hand



(a) (b) (c)

(d) (e)

Figure 3.21: The process of recognizing a touch event in PalmRC: (a) original depth image, (b) subtracted background, (c) determining the counter of the reference hand as well as (d) the palm box, and finally (e)recognized touch on the surface.

and recognize touch and dragging events with the index finger of the dominant hand. The built-in depth sensors recognize a user's hands in a minimum distance of 50cm. Currently, we mount the depth camera on a tripod located at the back of a user's shoulder (cf. figure 3.19). We envision future depth cameras to be small and precise enough to be either unobtrusively worn, or to be integrated into living room furniture.

Touch events are recognized in a multi-step process similar to [Gustafson 2011]. The process is depicted in figure 3.21. In order to subtract the background, we first find the closest pixels in the raw image and remove all other relative depth values greater than 40cm. We classify the depth values of each hand by calculating the number of peaks in a histogram of all depth values (cf. figure 3.21)(b)). To track the nondominant hand, we then calculate the contour and the convex hull of the hand including convexity defects (red points) and convexity start points (blue points) depicted in figure 3.21)(c). The palm box is then calculated based on the prominent defect and the start point (illustrated with yellow and light-blue circles in figure 3.21)(d) accordingly).

Touch events are recognized in a multistep process similar to [Gustafson 2011]. The process is depicted in figure 3.21. In order to subtract the background, we first find the closest pixels in the raw image and remove all other relative depth

values greater than 40cm. We classify the depth values of each hand by calculating the number of peaks in a histogram of all depth values (cf. figure 3.21)(b)). To track the nondominant hand, we then calculate the contour and the convex hull of the hand including convexity defects (red points) and convexity start points (blue points) depicted in figure 3.21)(c). The palm box is then calculated based on the prominent defect and the start point (illustrated with yellow and light-blue circles in figure 3.21)(d) accordingly).

To determine if and where the touch occurs, we compare the depth values of the finger tip with the surrounding values in the hand box. If the finger tip gets close enough to the reference hand, a touch event will be recognized. Due to the local noises and low-resolution of the Kinect depth camera, the precise end of finger tip is not fully recognizable. Similar to [Gustafson 2011], we determine the touch location by offsetting a small vector in the direction of the finger (yellow circle in figure 3.21)(e)).

Although the tracking approach requires users to hold their thumb upright while using PalmRC, it robustly recognizes different orientation of the nondominant hand. Future work is needed to improve the hand tracking and touch recognition so that users can arbitrary hold their hands.

Since, user input on today's televisions is supported through either conventional button-based remote controls or applications running on smart phones, we aim a third user study comparing these two well-established input modalities with PalmRC in terms of both performance and user experience. We present this comparative study in the following section.

3.4.5 Study3: Comparative Study

In this study, we identified user acceptance in a controlled laboratory evaluation comparing PalmRC concept with two most typical existing input modalities, here conventional remote control and touch-based remote control interfaces on smart phones for their user experience, task load, as well as overall preference. We focus on a set of basic tasks such as channel navigation and interaction with common TV applications.

3.4.5.1 Study Design and Tasks

We used a Samsung Smart TV and selected three different input conditions for our comparative study:

- the default button-based remote control for the utilized Samsung Smart TV used in our previous studies. We restricted the interaction to the directional keys and covered the rest of the keypad from the users to focus on basic TV interactions.

- an original Samsung Remote application (version 2.2.5) running on an android-based smart phone.
- PalmRC with the same setup as in the previous study. We connected PalmRC to the TV so that users were able to operate the original Samsung TV user interface. The study environment resembled a typical living room.

We used the original Samsung user interface for the tasks in all input conditions. The study consisted of three task sets. In task 1 and 2, participants were allowed to use four directional keys plus the OK button, as well as flick gestures for fast navigation. In the PalmRC condition the remote control mode with navigational keys was the only active interaction mode – no direct mapping of on-screen user interface was enabled. In task 3, participants could use directional keys, as well as special buttons such as play or pause in both remote control and smart phone conditions. Direct mapping mode of onscreen interface elements to the palm as region-based UIs was enabled in the PalmRC condition.

The participants had to complete the following tasks:

Task 1: The original list of TV programs (with a total of 43 programs – cf. figure 3.22(a)) was first shown on the TV screen. Participants had to navigate and find two specific programs. The programs were located at two positions in the list, one with a relative short navigational distance to the start position of the task, and the other with a larger navigational distance, respectively.

Task 2: In the second set of tasks, participants had to find and watch a movie trailer in the video-on-demand portal of Samsung's Smart TV interface (cf. figure 3.22(b)and (c)).

Task 3: In this task, we compared the direct mapping of on-screen items offered in the PalmRC condition with the common navigation techniques using directional keys of both button and smart phone based remote controls. This task helped us to compare direct selecting a target of TV UI elements with the common way of navigating to the target and selecting it. To do so, we used Shralp (cf. figure 3.22(d) and (f)), a snowboarding video podcast application due to the simplicity of its user interface. Participants had to first select a video from a

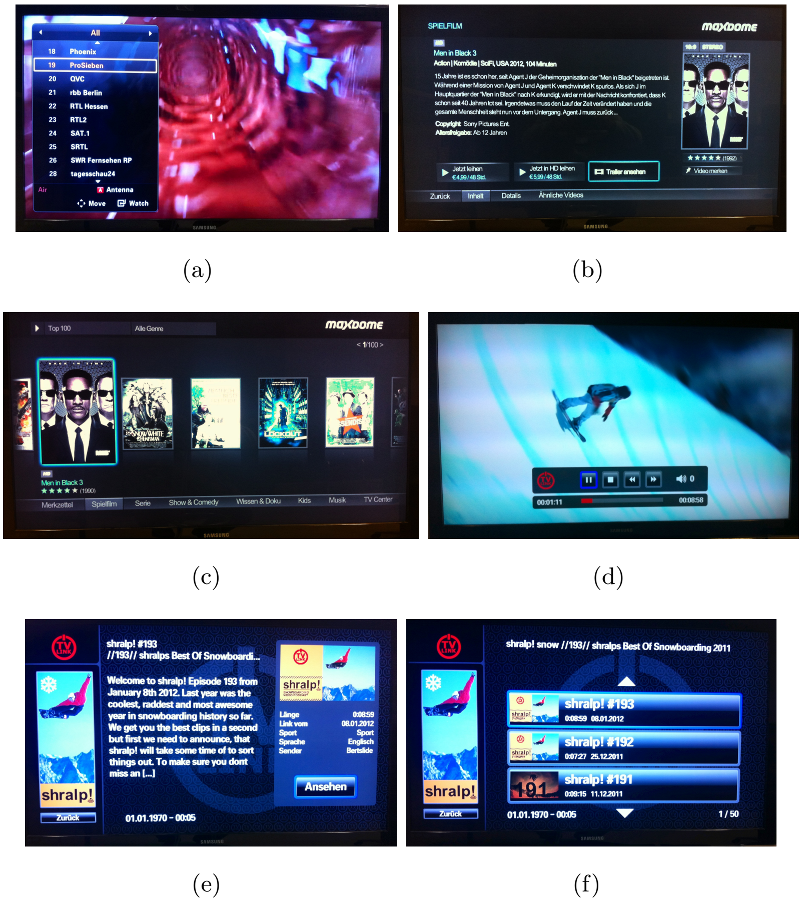


Figure 3.22: The Samsung user interfaces and applications used in the comparative user study

menu with four options. Once the video was played back, they had to seek for two positions in the video.

3.4.5.2 Methodology

We recruited twenty participants (three left-handed, seven female, 35 years of age in average) from various backgrounds such as household, school students, PhD and master students and administrative staff (like secretary). None of the participants took part in any of the previous studies. The participants were introduced to PalmRC concept and the prototype upfront. Each single-user session lasted about one hour.

The order of input conditions was fully counter-balanced. After each input condition (i.e., standard remote control, smart phone application and PalmRC), participants were asked to complete an AttrakDiff [Hassenzahl 2003] questionnaire, allowing us to measure (1) the user experience, and the (2) NASA task load index [Hart 1988] which estimates the cognitive demand. Post-study interviews were carried out to collect general comments on the participants' experience during the experiment. All sessions were videotaped. We analyzed the interactions, video recordings, salient quotes of the transcripts of the interviews and observational notes using an open coding approach [Strauss 1998].

3.4.5.3 Results

In this section, we first present the results of AttrakDiff and the task load scales followed by behavioral patterns which we derived from our observations. At the end, we discuss some common concerns raised by participants.

AttrakDiff
Figure 3.23 shows a user experience portfolio, which situates the three input conditions alongside two quality dimensions: hedonic (pleasure) and pragmatic (usability) qualities. The portfolio shows that PalmRC excels in terms of hedonic qualities. Its pragmatic qualities are comparable to those of the standard remote control. Overall, the portfolio shows a tendency for PalmRC toward being "desired". Among the three input conditions, it was perceived as the most attractive interface with a score of 4.5 on a 7-point Likert scale.

The Smartphone interface was perceived as mediocre in terms of both qualities and achieved the lowest score in terms of pragmatic qualities, hence its usability. In terms of attractiveness, it scored 3.5 points. The standard remote control was perceived as the interface with the lowest hedonic qualities, with a slight tendency toward being "too task-oriented". It was also evaluated as the least attractive interface with a score of 3. According to the participants, it was further perceived as "ordinary" and "unpleasant".

Task Load Index
We collected the perceived workload data using a scale of 1-20 (1 means the least effort) for various types of workloads; mental effort, physical effort, temporal demand, performance, overall effort and frustration. A 1-way repeated measure ANOVA found the interface (conditions) to have a main effect on physical effort, temporal demand and frustration (see figure 3.24). Post-hoc pair-wise comparison with Bonferroni correction revealed that the Smartphone interface caused significantly more

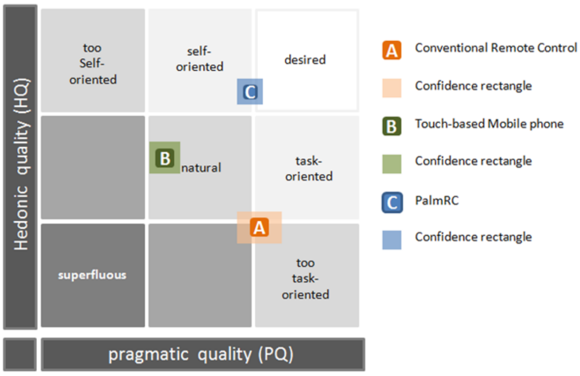


Figure 3.23: Portfplio with average values of the PQ and HQ dimensions and the respective confidence rectangle of each input conditions.

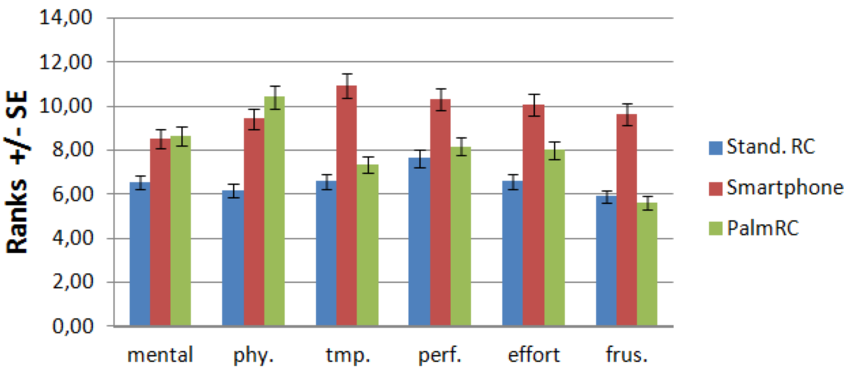


Figure 3.24: Portfolio with average values of the PQ and HQ dimensions and the respective confidence rectangle of each input conditions.

temporal demand (M= 10.9, SD= 4.1) and frustration (M=9.6, SD= 3.5) than the other two interfaces (both with $p<.001$).

In comparison to the Smartphone and PalmRC conditions, the standard remote control condition resulted in significantly less physical effort being reported (M=6.1, SD= 3.4, $p<.001$). We believe that this is because of the two-handed nature of PalmRC (and partially Smartphone) that may require more coordination of both hands.

Observations

General observation: We observed that almost all participants quickly picked up the interaction style of PalmRC. Participants found PalmRC intuitive, fun and entertaining. They appreciated the lack of visual feedback on the palm: "I just want to see the consequence of my action on the TV and that is enough". All participants agreed that PalmRC provides shortcuts and immediate interaction. Eight participants stated that PalmRC is very practical in situations where grabbing a conventional remote control or cell phone is difficult –e.g., when either of them are out of reach, or if the hands are soiled while eating something.

All participants found the standard remote control to be practical; for both simple (such as navigation) as well as advanced functions. However, due to its inadequate design it received minimal hedonic qualities. P4 stated "they [remote controls] are bad designed but well-used". Nine participants mentioned that they usually have to deal with more than one remote control because of different devices connected to the TV (such as satellite or audio equipments). In contrast, PalmRC was perceived by almost all participants as a universal and personalized input modality for different devices. Nevertheless, for more non-common interactions participants envisaged using standard remote control instead of PalmRC.

With respect to the smart phone interface, we observed that participants had more difficulties than with the standard remote control. This was mainly because of the high degree of attention required to look for virtual control widgets on the smart phone. We observed that the participants held the phone in their hands while watching the TV screen. In many cases this lead to unwanted touch events and accidental interactions. Eleven participants said that smart phones are more suitable as a secondary device where one can obtain additional information about the program. P4 said "I would like to check related tweets or my Facebook page while watching".

Visual attention: we analyzed how participants interacted with the interface in each condition. In response to R2 in section3.2.2, we were particularly interested to see how their visual attention was directed while using each input modality. All participants used PalmRC almost eyes-freely. They mainly preserved their attention to the television.

We observed a different behavior for the standard remote control condition. The interaction was basically performed in two phases. In the first rather short phase, participants grabbed the device and looked at it for a short while to get accustomed to the layout and find the most common keys (mainly directional keys). This behavior distracted the participants from the TV. In the second phase, participants left their thumb on one of the main keys (mainly on the OK button in the middle of directional keys) while their attention was directed to the TV. In this phase, interaction with the TV was performed almost eyes-freely for navigational tasks. Participants however mentioned that in order to look for a specific button, they needed to peek at the remote control.

Based on our observations, we found that the smart phone was the input modality which required the most visual attention. Participants needed to look at the display of the smart phone for nearly every single interaction. (1) Lack of an overview of all functions and, (2) need to switch to different application mode for different functionalities were the main reasons that participants pointed out.

Concerns: the different orientations designed in PalmRC (cf. figure 3.11) turned out to be confusing for eight participants. The participants had particular difficulties with mapping four directions to their hands. Discussion with participants revealed that directional keys should always be oriented toward the TV. This means that the upper part of the hand surface should be mapped to the UP key, the lower part to the DOWN key respectively.

Nine participants were concerned with the required two-handed usage of PalmRC in contrast to the one-handed usage of the standard remote controls. P4 commented, "With my remote control at home, I can control the TV while I'm holding a glass in my other hand". We believe that this issue becomes less severe by extending the PalmRC concept to surfaces of other body parts such as thighs, which affords one-handed interaction. Moreover, the Kinect depth sensing technology has opened up new interactive experiences leveraging any un-instrumented physical surface around users such as couch arms or tables as an input surface to operate TVs.

3.4.5.4 Summary, Discussion and Limitations

The results from comparative study are summarized in table 3.8. Overall, we found that PalmRC provides a usable and foremost joyful way for TV remote interaction. Our observations suggest that is mainly due to its touch-based, eyes-free input characteristic, as well as the natural haptic feedback provided through one's own body parts. It is important to note that PalmRC is not meant as an alternative, but a complementary input technique for TV remote interaction. The study revealed that PalmRC provides a shortcut for common TV interactions and therefore can improve the overall user experience while watching TV.

Furthermore, the study findings confirm existing assumptions that smart phones and other "secondary interactive screens" in the living room are more suitable as a companion device than a "first-class device" demanding a user's complete focus.

While most findings show that PalmRC can enrich the overall living room TV experiences, some limitations apply due to the setting of the study. Although we aimed to create a more realistic environment, the study neglects contextual influences of the real life living room. Among the others, various living room arrangements, num-

Evaluation Results
– PalmRC provides a joyful way for TV remote interaction as it is eyes-free and provides natural haptic feedback.
– PalmRC is not meant as an alternative, but a complementary input technique.
– PalmRC provides a shortcut for common TV interactions.
– Various living room arrangements, number of viewers, and their postures in front of the TV as well as their age and health abilities might effect the usability of PalmRC.
– Interactions proposed by PalmRC are still basic and simple compared to the other input mechanisms.

Table 3.8: Summary of the main findings from PalmRC controlled experiment study.

ber of viewers and their postures in front of the TV as well as their age and health abilities are instances that are not considered in the study. On the other hand, the novelty of the PalmRC concept might influenced why participants rated PalmRC as the most desired input mechanism for TV interactions (cf. figure 3.24). While this limits us in investigating the natural user experience in living rooms, the study revealed the salient characteristics of each input mechanism and how they can best complement each other.

As another limitation, the current implementation of PalmRC offers a set of limited interactions with TV systems. It basically supports shortcuts for a few simple functions as well as nonlinear navigation and direct selection of UI elements. As a result, the evaluation focused on studying a set of basic and simple interactions compared to the other input mechanisms. While we believe that this is valid and important as the first step, more advanced interactions need to be examined.

3.5 Conclusion

In the present chapter, we bring the body-based spatial interactions to interactive televisions at homes. Unlike traditional TV interfaces that are either device-based (e.g., remote control) or visually demanding (e.g., touch-based screens), body-based user interfaces are deviceless, omni-present, and eyes-free. Such interfaces have a high potential to foster intuitive experiences. In addition, despite existing eyes-free TV input modalities (e.g., voice and gestures), body-based interactions have the ability to support implicit interactions using the whole body postural information. We also investigate leveraging spatial information of a part of body, particularly palm surface, to support more fine-grained TV interactions. We argue that using the palm surface as an imaginary eyes-free remote control opens up a wide range of new interaction possibilities between television and viewers.

In this chapter, we contribute two body-based TV user interfaces, namely CouchTV and PalmRC, that alleviate course- and fine-grained interaction techniques, respectively. Based on these, we showcase different TV application scenarios. The design of both interfaces is grounded from an extensive literature analysis of body-based user interfaces and by results of a preliminary study presented in Section 3.2.1 that aimed to better understand people's use of their bodies in front of the TV.

CouchTV relies on spatial and postural information of viewers, such as user's presence, location, orientation, and pose. We design novel interaction techniques to support (re)engaging in TV watching activity and provide appropriate level of awareness. PalmRC demonstrates the concept of leveraging the palm surface as an eyes-free remote control. It allows TV viewers to either interact directly with UI components or map remote control functionalities to the surface of the palm.We protoypically realize both concepts CouchTV and PalmRC, have to validate their design and interaction concepts and conducted a series of studies.

In early user feedback, we investigated the CouchTV usage and its watching experience attractiveness. There was a strong consensus that CouchTV is intuitive and practical as it optimally assists and motivates the users. Since it reacts to viewer's spatial presence and postural movements, it resembles the sense of having an ambient display at home. Users also found it as a lightweight means for obtaining information about TV contents while watching.

We also observed participants raised concerns about the design of the CouchTV interface, for example, regarding the synchronization between the TV reactions and spatial situations of the user. Therefore, being in sync and the immediacy of body-based TV responses is a must. We also found that CouchTV induces a limpness between coviewers (i.e., the more they received content awareness through CouchTV, the more they were disconnected from coviewers). Nevertheless, the actual disconnection is not to be regarded as something negative, since intentional disconnection may also enhance the overall experience.

The foundations of the PalmRC interface were grounded in an exploratory study. We gained qualitative insights into how people would use their hand as if it were a remote control. Results suggested that users are able to touch several salient regions of their palm without looking at them. In a controlled experiment, we quantitatively determined how precisely they could interact with these regions in an eyes-free manner. We also investigated the effectiveness of using the palm as an input surface for direct interaction with on-screen user interface elements.

The findings showed that under certain circumstances (e.g., 28 mm button size and four target options) the palm-based remote control is viable. In the third study, being comparative in nature, we contrasted PalmRC with two common TV input modalities: standard remote controls and smartphones. The results shed light on advantages and disadvantages of each input modality in terms of both usability and user experience. The results further underline the fact that PalmRC offers an always available, efficient and effective shortcut for performing frequently used interactions with TV systems without requiring additional mediator devices.

We conclude that by leveraging different landmarks of the hand, users are able to perform precise interactions, while preserving their visual attention to the TV. At the same time, the palm surface is also appropriate for gestures, such as swiping. Based on the initial results, we hypothesize that the interaction style of PalmRC is less tiring than mid-air gestures. Future studies are needed to systematically compare both as deviceless input modalities for TV systems.

Chapter 4

Connecting Shared Event Experiences

Contents

The last two chapters aimed at designing, developing and evaluating novel appropriate interaction concepts for watching experiences in two distinct settings to support both in-site spectators and remote viewers at homes during live events. Chapter 2 targeted in situ experiences during local-scope mass events. We proposed live sharing of user-generated videos in real time to support quick access to different viewing angles of the event. We further investigate how this has the potential to socially support the co-construction of the overall event experience where attendees share the same event and the same location. In Chapter 3, we then explored the TV watching experience in *living room settings*. We showed that how body-based spatial interfaces can go beyond the remote control paradigm to enhance the interaction with social televisions while one or several event fans (viewers) gathering together in front of a television set to watch television. The proposed concepts also opened up a wide range of novel interaction possibilities between viewers, the television, and contents, which are less distracting in case of following a live event when viewers are not willing to miss even a single moment of it.

Although experiences in both settings concern the same event, they are fundamentally different. People in the field, who witnessed an event through both listening to the atmosphere and peripheral vision, perceive the event differently than those people in living rooms who watch the event with the limited perspective of professional broadcasts. People in living rooms may have access to the additional information such as an audio commentary, scene replays, or even information from the Internet (such as an event's Tweets). However, they lack (1) limited viewing perspectives, (2) social interactions with the attendees in the event, and (3) sharing event experiences in the "opposite" realm (e.g., people at home cannot contribute expressions such as emotions to the event experience in situ and vice versa.).

The present chapter connects the experiences of these two settings and at the same time transgresses the limitation of each. We go beyond the concept of CoStream presented in Chapter 2 by supporting the bidirectional communication channel from in situ sharing toward remote sharing with viewers at home and thus, bridging the gap between the two settings. We started by conducting a preliminary study to deepen our understanding about people current communication and social patterns in front of the TV. We were especially concerned with identifying how people typically communicate with noncollocated viewers (who are either watching in another home settings or in the field). More precisely, we drilled down through interpersonal relationships and investigated how this can be considered in the design of concepts aiming to connect people at live events and those engaging in front of TV.

The study findings showed that the nature of some relationships can be matched to certain genre characteristics. Different social relationships remarkably demonstrated specific behavior related to the social watching experience. It was found that sports genre was mostly preferred to be watched in social groups. People also preferred to watch sports almost equally with remote peers regardless of their social relationships. This highlighted that in the sports genre, the quantity of noncollocated coviewers (particularly those that are in the fields) is more important than the social pattern preferences.

Based on the results of this study, we then focused on bridging the experiential gap existing between between people in the field and those engaging remotely during *live sporting events*. To address this, in this chapter we contribute a set of interaction concepts and techniques to connect both types of spectators through bi-directional mobile live video sharing, which we call *CoStream@Home*. We exploit mobile devices in both realms as a means for mutually contributing to the event engagement, potentially leading to more immersive and socially connected experiences during such events.

In order to stimulate social interactions, we consider establishing real-time communication and gestural information of viewers in front of the TV (e.g., emotional and gestural reactions) in addition to the video sharing communication. We illustrate how we envision this to aid in bridging the aforementioned gap and believe that such information can open up novel social interaction appropriate for both realms. The effectiveness of CoStream@Home and interaction techniques is tested in an early user feedback session.

In summary, contributions of this chapter are the following:

- The preliminary study identifying social patterns while watching TV
- Design of CoStream@Home concept along with a set of interaction concepts
- Evaluation of the concept through an early user feedback session
- Design implications for future connected TV user interfaces

The remainder of this chapter is organized as follows. In Section 4.1, we begin by presenting an overview over previous research addressing the connected user experience between the field and people watching live professional broadcasts in the living room. Section 4.2.1 presents the preliminary user study in which we specify social patterns based on genres and viewers' interpersonal relationships while watching TV.

We then describe our concept namely, CoStream@Home, and the underlying interaction techniques and its technical implementation detailed in Section 4.3. Later in Section 4.4, we report on an evaluation in which both the system concept and the user interface are examined. We further provide a set of generic design guideline for future TV user interfaces supporting remote watching experiences. At the end, this chapter concludes with a discussion presented in 4.5.

Contribution Statement: Most of the work presented here is based on [Dezfuli 2011, Dezfuli 013, Dezfuli 2013a]. I am the first author of these publications. I have initiated and lead the project. My coauthors have also contributed significantly. Master students, Florian Müller, and Sebastian Günther, have built and implemented many aspects of the CoStream@Home system. My supervisors, Jochen Huber, Mohammadreza Khalilbeigi, and Max Mühlhäuser, have contributed to the design of the system and in writing the papers.

4.1 Related Work

Enhancing the social connectedness between two realms, namely homes and field, has recently drawn the attention of both Multimedia and CHI researchers. In this section, we review the state-of-the-art that focused on the social connection and interactions among these two realms. We start by discussing prior research aimed at understanding important factors that are unique to the field and are missing while watching from home and vice versa. We then present prior systems that support bridging the gap between the field and home through sharing real-time video, emotions, and gestures.

In Section 4.1.5, we moreover discuss previous work connecting noncollocated TV viewers at different households for active participation in the live event. Although the focus of this chapter is on connecting the home and field, we believe that reviewing this stream of research can provide valuable insights into the design of our system. A short summary of the literature examination is given at the end of this section.

4.1.1 Event Participation

There are various studies that focused on identifying special characteristics of following an event and the most important reasons why people attend events live or follow it from home in front of the TV [Uhrich 2010, Trail 2005, Wann 1996].

Field

Uhrich and Benkenstein [Uhrich 2010] addressed the concept of atmosphere through a multistage expert survey. They found that the organizers of an event, spectators, actions of a match, and stadium architecture are key environmental factors that create a unique atmosphere in stadium-based events. Furthermore, based on the study results, the authors argued that these factors lead to amazing vibes, tremendous enthusiasm, strong emotions, and intense euphoria while experiencing

Figure 4.1: Field studies conducted in sport stadiums in shanghai [Sun 2007].

the event live.

Despite all its unique atmosphere,following an event in the field is still not a perfect experience. Several prior work addressed the fact that the spectatorship experiences of live events have some deficiencies, such as lack of detail about ongoing event, that may diminish the user experience. Sun and May [Sun 2007] conducted a set of user studies at two swimming galas and two football matches examining the user experience of the event in the field. The studies revealed that current user experience at the large sport events was not considered as a positive experience and people expressed a higher expectation when in the field. This is related to lack of detailed information and social interactions. They found that unlike TV watching at home, spectating in the field lacks the detailed information about the ongoing event. The spectators detailed information was limited the stadium loudspeaker system and large screens, which are typically located at the hot spots of the events where many spectators were far from. Moreover, the broadcasted information on such screens were not under the user's control and were only partly relevant.

It was also found that spectators spent seldom or only a small portion of time of the event duration for social interactions. The crowds imposed difficulties in meeting friends at the stadium while or even after the events. As a consequence, spectators staged their experiences mainly by watching, and they tried to move around the stadium to optimize their viewing angles.

Home

On one hand, due to the limited capacity of arenas, physical distance, and cost issues, a large number of fans follow the event remotely at homes through professional broadcasts. Potentially, this can result into a limited experience that is bounded to the virtual world of televisions and ignores the users' emotional responses to the ongoing events' actions. Consequently, TV viewers at homes may feel less connected and more lonely, compared to when being in site [Raney 2006]. On the other hand, there are studies showing that there are numerous factors that potentially contribute to the enjoyment of sports presented on television. Bryant et al. [Bryant 2000] compared the physical participating at sporting events versus watching sports on television screen and found the motivation behind watching sport event at home is due to *"plethora of sports contests on the screen."*

Given the pros and cons of both realms, there is a number of studies that aim at addressing this bi-directional problem by bridging the physical distance between homes and the field. These studies explored if the field experience can be enhanced through content "snacking" that is sharing and posting contents (e.g., tweets or messages) through handheld mobile devices. We categorize the studies in this vein of research into three groups based on the type of content: 1) user-generated videos, 2) emotional, and 3) gestures. These are discussed in the next section.

4.1.2 Video-Based Communication

Mobile user-generated video broadcasting has been widely used as a means to connect live event and homes [Juhlin 2010, Engström 2012, Kaheel 2009].

Juhlin et al. [Juhlin 2010] investigated the use of mobile broadcasting services as a social medium shared between mobile users and remote viewers over distance. They particularly focused on the topics people are interested to share via user generated live video and how they represent those visually in this media format. The available postings of four popular websites (e.g., bambuser.com) were examined. The authors found that apart from technology tests videos in which the users just tried to become familiar with such services, people mainly broadcast to feature social events of various kind where people, groups, and crowds seem to be the main topic. They also observed that people choose an interview format, such as turning camera to himself and the people present, and talk about ongoing actions around. This was found as a format that implicitly coordinates the live appearance and activities in front of the camera. This study is an instant evident that mobile broadcasting is a common social medium whose users leverages its specific affordances to connect remote people in shared experiences.

Similarly, Engström et al. [Engström 2012] proposed a mobile system, namely Instant Broadcasting System (IBS), that allows users to collaboratively create user-generated live video contents from multiple networked camera phones. They can edit video feeds into a more visually interesting story in real time. IBS is particularly designed for amateur camera users who attend large events for their own enjoyment. It consists of client and server applications. The client captures, decodes and transfers video streams. The server application is basically a video mixer user interface that displays the incoming video streams. The authors evaluated the system through a field trial that was conducted in a music festival where the users were able to share their own generated content on public screens. The study showed that collaboration for producing one compound visual story out of individual camera feeds is very challenging as it needs a communication back-channel. While text messaging was found to be an attentive communication means, audio chat was only possible in one direction as the camera operators cannot talk while filming and it will be useless if the number of operators scales up.

Mobicast [Kaheel 2009] is another system that was designed for collaborative live video streaming. The system enables attendees of an event streaming video from their mobile phones and makes a collective viewing experience of the event for remote viewers. The collective view is supported either through the selection of best live viewing camera perspective or stitching the individual live feeds together, which presents a wider field of view (e.g., a panoramic video of an event). Accordingly, the users could mutually receive feedback from other streamers about how well their videos fit (or stitch) with other video streams.

The performance of the system was evaluated through videos streamed during field-trials. The findings showed that the system performance was quite satisfactory and the proposed feedback was estimated as a very useful feature that could increase the system precision. Therefore, the proposed approach can be seen as a promising way for visualizing user-generated live video during events in order to enhance the viewing experience in a social manner during events.

Guimaraes et al. studied social practices around ultra-personal videos within groups of people with strong ties such as friends and family [Laiola Guimarães 2011a]. They conducted a user study in the context of a small-scale school concert where performers and participants belong to the same social circle. It was found that users appreciate the importance of video sharing for feeling more connected and with other group members and building shared experiences. Their results also implied that future video sharing systems should not only provide useful mechanism for navigating and sharing complex media information but also take into account the emotional intensity and intimacy factors. Inspiring by these findings, later in this chapter we also study the impact of fine-grained interpersonal relationships around various TV genres with a particular emphasize on live event programs such as sporting genres.

Photo by Webb Chappell, copyright 1999 Webb Chappell; reprinted with permission

Figure 4.2: Self-report application: the user taps the thumbs-up after getting a cup of cappuccino [Picard 2000].

4.1.3 Emotion-Based Communication

In addition to prior research focused on live user-generated video broadcasting, there are a few approaches that proposed sensing human emotions as a communication channel between in-venue fans and remote event viewers [De Silva 1997, Picard 2000, Russell 1989].

De Silva et al. [De Silva 1997] conducted an evaluation experiment to test the human beings' recognition considering six emotion sets (i.e., angry, happy, sad, surprise, dislike, and fear) using either audio information, video information, or a combination of both. The study results showed that anger, happiness, surprise, and dislike are four emotions that are highly visually recognizable. However, sadness and fear are better recognizable by audio information.

Picard [Picard 2000] aimed at enabling computers to sense, recognize, and respond to the human emotion in order to communicate more naturally and implicitly. He developed two systems, namely, self-report and concurrent expression, that can intelligently handle common human expressions such as frustration, confusion, and boredom. Self-report is a system that allows users to communicate either by means of software (e.g., a menu of emotions with words or icons) or by touching hardware (e.g., tangible icons) and let the system know about how users feel. This is shown in figure 4.2. In concurrent expression, the system can sense the user affective expres-

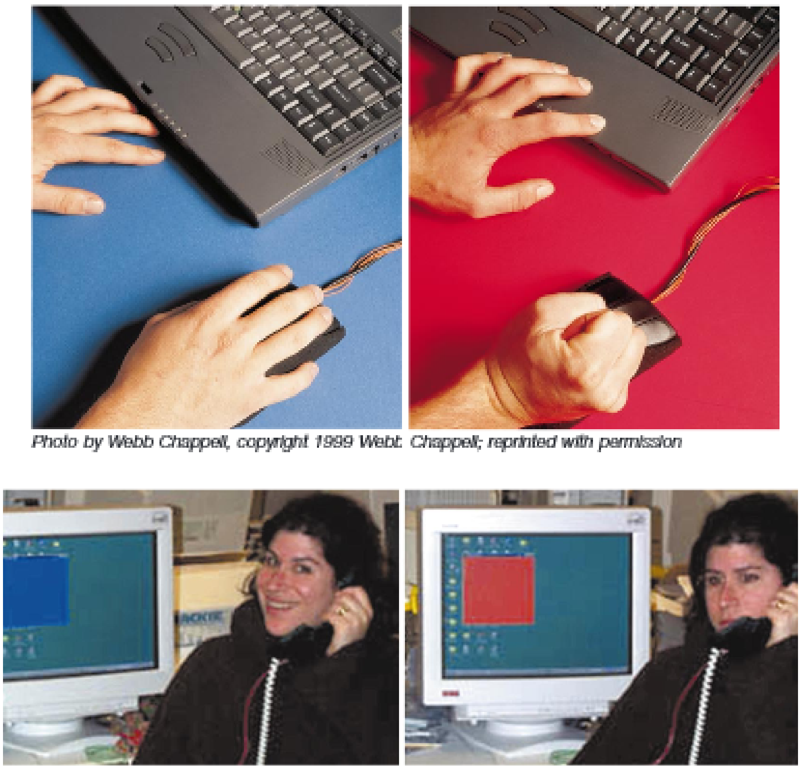


Figure 4.3: Emotional intelligent systems using pressure sensors [Picard 2000].

sion in a much more natural way using different sensors (e.g., microphone or pressure sensor embedded in a mouse or phone as it is shown in figure 4.3). He stated that the ideas behind such applications raise the potential to design future machines that are intelligent in responding people's emotional expressions in an appropriate way.

Russel [Russell 1989] introduced a system, namely Affect Grid, to enable describing human mood and feeling based on the facial expression along the dimension of pleasure-displeasure and arousal-sleepiness. The system is a 10x10 table, where users were required to check the grid position that best describes their current disposition. The Affect Grid was evaluated through four user studies and was acknowledged as an adequate reliable and valid tool for measuring the mood. This can be used for future systems that induce the emotions experienced by the fans in the field, to examine if remote fans have similar emotions while watching at home.

4.1.4 Gestures-Based Communication

Finally, gesturing in front of television sets has caught the attention of few researchers to enhance the viewing experiences and propose more emotional and social interactions about live events [Juhlin 2013, Centieiro 2013].

Juhlin and Önnevall et al. [Juhlin 2013] analyzed natural physical movements of viewers in front of TV screens collected through an ethnographic observation in sports bars and private living rooms. The study ran during Winter Olympic Games in Vancouver and focused on identifying relations between viewers' gestures, the group of collocate viewers, and the event displayed on the screen. Their results reflected that there is a set of gestures commonly occurred in front of TV -(e.g., lifting the arm to cheer, covering faces with hands, etc.) that can be useful for the design of natural gestural interaction techniques in collaborative settings. They also added that the gestures are continues and negotiated and can be influenced by the ongoing action on the broadcast or other viewers in the group.

Centieiro [Centieiro 2013] proposed two novel mobile prototypes, namely, WeApplaud and WeBet, that can bring the venue atmosphere, its immersion, and emotional experiences to remote viewers while watching soccer matches live on the television (cf, figure 4.4. WeApplaud is a mobile game that enables multicollocated players to be engaged with the applause happening in the stadium. The system leverages the notion of award to encourage the viewers to clap like they do in real life using their mobile phones and synchronized with the audience in the field. WeBet is the second prototype that allows users to bet if a goal is about to happen using their mobile phones in an eyes-free manner. Both system prototypes were evaluated in an early user feedback. The analysis of the results revealed that the proposed systems make the remote watching soccer matches more interesting and increased the level of entertainment and joyfulness.

In summary, the previous work discussed above highlighted the pros and cons of experiencing the event at homes and in the field and addressed the gap between these two realms. Researchers proposed that the future systems can largely benefit from sharing of user-generated videos, emotions, and gestures for bridging the experimental gap. Later in this chapter, we show how we leveraged the findings from this stream of research in the design of CoStream@Home to support the coexperiencing live events for connecting fans located at homes and in the field.

4.1.5 TV Viewer's Communication

This section reviews the state-of-the-art focusing on communication over TV broadcast within multiple households to offer rich social experiences. There exist a large body of work that investigate how TV viewers in different locations can communicate with each other using various communication modalities (e.g., text, voice, video

Figure 4.4: WeApplaud and WeBet Interactions supported for remote fans [Centieiro 2013].

chat, etc.) [Harboe 2008, Coppens 2004, Geerts 2006a].

Bernhaupt et al. [Bernhaupt 2008] explored the ongoing trends around television in the home environment in two ethnographic studies. For the studies, they developed two types of cultural probes method including creative and playful probing to become more creative and to minimize the negative effects of researchers taking part in field studies [Gaver 1999]. Here, they modified a card game by adding research-related query cards to enhance the participants' involvement in the study.

The results showed that there is a need for ubiquitous connectivity and communication with others over a distance to support future living rooms expectations. They further argued that extended home, the shared experiences and new interaction techniques for televisions are the three highlighted trends in home context. While these trends were clearly desired, it was concluded that some challenges like privacy and security issues, have to be considered in the design of new interaction techniques.

Geerts and Groof [Geerts 2009] presented a set of sociability heuristics for de-

Figure 4.5: Media sharing interface of the system used in the user study [Geerts 2008b].

signing or evaluating social interactive televisions. Based on the analysis of various social television systems and the results of several user studies, they stated that designers of such systems should minimize the distraction from the television program while at the same time encouraging shared activity both with collocated and remote viewers. Furthermore, the authors argued that social television systems need to offer different synchronous and asynchronous channels (e.g., text and audio) and levels (e.g., emoticons and automatic replies) for free-form communication between households. It was also shown that social TV is not only about communication and sharing user-generated content, such as some form of commentary or a user's cut of an interesting video, but also, they should support establishing successful social connections in front of TVs between homes.

Weisz et al. [Weisz 2007] explored the potential of chatting while video watching to engage remote viewers in an active social social experience. They conducted two exploratory studies in the laboratory to first investigate if and how people can be distracted by text chat while watching a movie. They found that watching videos, in particular live broadcast, can end in frustration rather than a desired social experience of watching together. Based on these findings, they proposed and examined two different method to reduce the imposed distraction: (1) merging the video content with several intermissions, and (2) enabling the discussion only after the video. The study revealed that chatting has a positive influence on social relationships between friends and even strangers. They also found that the two proposed methods can potentially provide solutions for reducing distraction but are not appropriate for all types of genres.

Geerts et al. [Geerts 2008b] focused on how television genres can play a role in the use of social interactive televisions. They designed and developed a system that enables talking and sharing audiovisual content between users (e.g., sharing

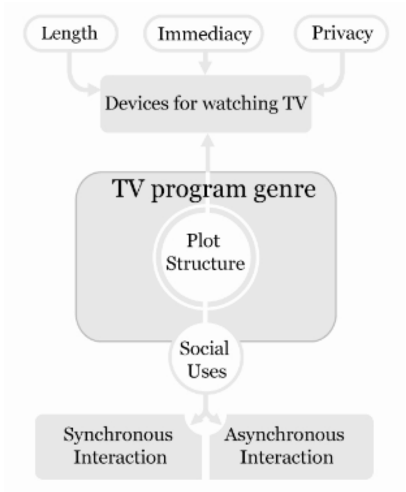


Figure 4.6: The role of genres on people communication. [Geerts 2008b].

annotated video clips to friends or family as it is shown in figure 4.5). They used this system to perform a user study in a simulated living room to find out the effect of genres on communication while watching TV.

The study results showed that the density of plot structure (i.e., events that make up a story) in a genre can directly affect communication while watching. However, the preferences in talking and sharing activities while watching different genres are a very complex issue and can be dependent on other factors (e.g., social habits and devices in which people are receiving videos – cf. figure 4.6). Based on their findings, the authors concluded that the design of social iTV systems should take into account this interplay between these social uses and genres to offer a rich watching experience to TV viewers.

Overall, the review of the estate-of-the-art above revealed that researchers have widely studied supporting various forms of synchronous and asynchronous remote communications between noncollocated viewers. While the main focus of this stream of research is on the communication features, little has been explored on identifying the relation between genres and people's interpersonal relationships and its effect on shared experiences while watching TV. Therefore, in a preliminary study presented in Section 4.2.1, we go beyond what has been previously studied by investigating the role of interpersonal relationships while watching and its link to video genres.

4.1.6 Summary

In summary, we reported on prior studies that aimed at bridging the experiential gap between viewers at home and people in the field. The studies mainly emphasized on drawing people into a shared experiences and live contents to serve a basis for socialization and enhance the event experience. It was showed that mobile video broadcasting, emotional intelligence, and gesture-based communication can potentially provide solutions for connecting people in both settings, effectively.

We also revealed that the prior work mainly focused on connecting noncollocated TV viewers on different household via telecommunication technologies for active participation in live event. These technologies included presence of viewers, text and audio chat. While this generates a potential to create remotely shared experiences around TV content, it may also inflict the feeling of distraction from actual TV program content as the main research challenge.

The state-of-the-art found that investigating the TV genre may help better explain this challenge as it play an important role in the use of social television, especially regarding attention and communication activity. However, these findings are based on lab studies and self-report questionnaires rather than longitudinal study. This mystification motivated our preliminary study presented in next section. It investigates the relation between different genres and interpersonal relationship and their effect on communication while watching experiences. The results of the study led us to certain guidelines that are supported in the design of CoStream@Home: a set of interaction and communication concepts to bridge the gap between homes and the field.

Table 4.1 summarizes the main findings derived from the related work.

4.2 Understanding Social Remote TV Watching

4.2.1 Preliminary Diary Study

As a response to the need for social interaction, researchers have been working on STV, a popular and already social medium, with social features enabling remote interactions, which refers. As shown in the related work section, these systems, however, do not take into account the variation in relationships people have. People have diverse interpersonal relationships as they are participating in various communities, which make them belong to various groups of friends, each of them with different common interests. Therefore, a buddy list is not a group of equal friends as people's relationships have diverse and complex structures [Spencer 2006].

We believe that depending on the video content, social television has a greater potential to provide feelings of togetherness if real-life relationships are taken into account. In addition, as the state-of-the-art revealed, genre can play an important

Findings
– The spectators, actions of a match, and stadium architecture are key factors that create unique atmosphere in events.
– Spectatorship experiences cause intense euphoria but lack the detailed information about the event.
– The viewing experience at home is somewhat limited to a virtual world trapped inside the television.
– Sharing user-generated videos, emotions, and gestures can help connecting people between the home and field.
– The prior work mainly worked on supporting various forms of sync or async remote communications between noncollocated viewers.
– The effect of interpersonal relationships and its relation to genres on remote communication is less explored.

Table 4.1: Summary of the state-of-the-art analysis. This summary underlines the requirements that are addressed with the contributions of this chapter.

role on social communication while watching noncollocatedly. Therefore, we first aimed to systematically explore the social interpersonal relationship pattern to see how and in which genres social watching is preferred.

In this section, we report on a preliminary study of people's video viewing habits (broadcast or online), targeting both the current and preferred social context of viewing. Our main objective of the study was to investigate the preferred social structure of watching video. Specifically, we studied the social context based on a fine-granular classification system of interpersonal relationships. This led us to derive some patterns that were elaborated by analyzing the relation with video genres. With this study, we wanted to answer the following research questions:

- What are preferences for social connectedness while remotely watching video materials?
- What is the relation between genres and interpersonal relationships?

People have different classifications of their social circle. In order to have consistent data, we created a comprehensible classification of people's relationships including family members, friends, strangers and "nobody," referring to a non-social situation (see table 4.2). Moreover, we break down the friendship variations into six categories ranging from confidant to associate friends. These are highly inspired by

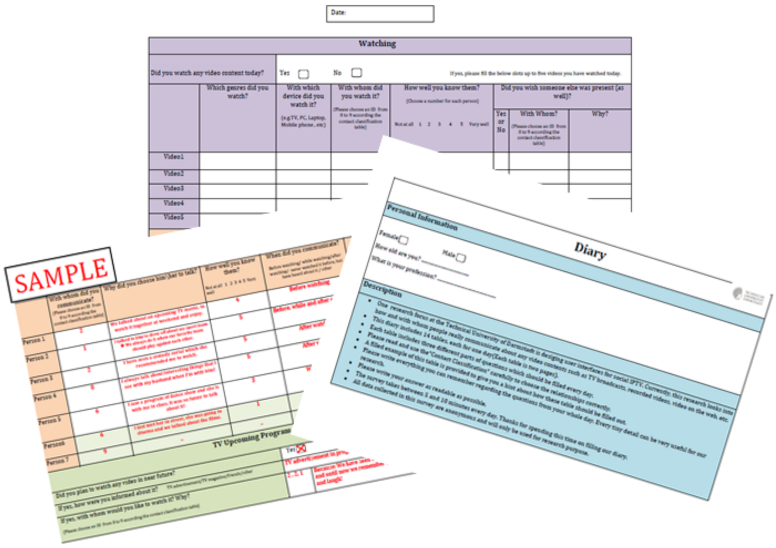


Figure 4.7: One diary and one sample of a completed diary were provided to participants to gain insight into people's social watching preferences.

Spencer and Pahl's work as their findings are based on strong empirical materials and fully cover a wide spectrum of friendship repertoires [Spencer 2006]. To ensure that the categories are comprehensible and distinguishable from each other, we performed a pilot study in a small scale. This resulted in a fine-grained yet simplified classification of relationships to help participants better categorize their social circle ranging from strong to weak ties based on the time, the emotional intensity, the mutual confiding, and the reciprocal services [Granovetter 1973, Gilbert 2009].

4.2.1.1 Method

We opted for structured diaries as an indirect observation technique, because it allows us to get a deeper insight in people's behavior of watching video over a longer period of time. The diaries were carefully designed to collect as detailed information as possible about watching habits. The participants had to describe in detail which videos they watched each day and how. We designed the main body of the diary in a way that allowed participants to elaborate more on the videos with respect to different questions. The questions provided us more detail about situations in which the participants watched (either together or alone, with whom on which device) as well as genres of videos.

Moreover, we asked participants whether the current situations fulfilled their social needs or not. If not, we asked them to detail their preferred relationships

Figure 4.8: Participants used our diary everyday while watching any video content.

based on table 4.2 and briefly describe the reason. We also provided a sample of a completed diary that gave the participants the opportunity to see how they should fill out the diaries (cf. figure 4.7 and 4.8).

The study was conducted in Germany and South Korea with an equal number of participants living in household types from families without children (13) to families with one (3) or two children (6). We recruited 22 participants (each country has 11 participants) for the study that lasted two weeks. The age of participants ranged from 23 to 50 years, with an average age of 30 (SD=6.80) and 38 (SD=5.62) in Germany and Korea, respectively. Nine of the participants were female and thirteen of the participants were male.

The participants were selected on the basis that they live in a household and watch live TV or on-demand content at least several times a day. We also ensured that the participants could assign their current relationships with others to all categories according to table 4.2. The majority of the participants in both countries had university degrees, and in terms of occupational classification, 14 people were considered as professional or managerial (such as senior research scientist and engineer), three participants were administrative (such as secretary), and five were housewives. We thanked all the participants with 10-euro Amazon gift cards.

4.2.1.2 Results

In total, the number of video samples elevated to 488 for the whole study (Germany (G):250, Korea (K):238 videos). Each participant watched an average of 22 (SD=3.33) and 21 (SD=1.49) videos in Germany and Korea, respectively. Further-

Relationships	Description
Close family	Family members who live together.
Far family	Family members who do not live together.
Confidant	Intimate friends who disclose their personal information and enjoy each other's company.
Comforters	People who help each other not only in a practical way but also give each other emotional support.
Fun	People who socialize together, but only for fun. They do not support each other with a deep level of emotional support.
Favor	People who help each other only in a functional manner.
Useful contact	People who share information related to work or advancing ones career.
Associate	People who only do or share a common activity like a hobby.
Stranger	People who do not know each other at all.

Table 4.2: Classification of interpersonal relationships

more, each participant also reported having watched four to six different genres. In terms of device usage, participants watched videos with the same types of devices in both countries: TV (G:64%, K: 68% of videos), PC (G:14%, K:10% of videos), laptop (G:14%, K:5% of videos), and mobile phone (G:8% , K:17% of videos).

Due to the differences in social and cultural norms across countries, we analyzed the data of each country individually. We performed a qualitative analysis, building up and iteratively refining our coding scheme during the analysis of the diaries. Social and non-social situations were coded separately for relevant behavior. For each country, we first describe the current social context of video watching. This is followed by participants' patterns of their preferences for social connectedness that are further elaborated with the type of genres. Finally, we describe non-social patterns.

Germany

Current Patterns of Social Watching: 112 videos recorded in the diaries from German participants in which they were collocatedly viewed in social situations (44%

of videos). 60% of the whole videos were watched with strong ties such as close and far family members as well as confidant friends. This can be partially due to the social context of the participants, living alone or in households as seven participants pointed out the fact that they watch more together to spend more time together.

For the rest of the categories (which encompass mainly weak relationships), the number of videos considerably decreased except associate friends. Participants mainly reported in diaries that they assigned their colleagues to this category. *"I'm on a business trip with my colleagues, we do everything together here*." For 5% of the videos viewed in Germany, four participants also reported that they watched these videos with strangers on their cell phones and laptops. *"He [a stranger] asked me about what I was watching and laughing at on my mobile phone*."

Preferences of Social Connectedness: 172 videos documented in Germany (68% of videos) were preferred to be watched in social situations. In contrast to the established (current collocated) social viewing practices that favored strong ties (family members), the desired (collocated and remote) future viewing practices favored weak ties. Nearly 50% of the videos were preferred to be watched with confidant, fun, and associate friends.

Confidant friends (31% of the whole videos, all participants) was the most preferred category to watch videos with. A reason might be that confidant friends not only support intimacy and trust like family members, but also, people in this category involve a high level of enjoyment [Spencer 2006].

Participants strongly preferred fun friends to watch soaps and comedy series (37% of soap and comedy videos, 10P). This can be explained by the fact that this is a category that revolves mainly around having fun [Spencer 2006]. They reported that increasing the joy of the watching activity (9P) and repeating funny statements with friends as the main reasons for that (6P): *"He [the participant's fun friend] made me laugh more"* or *"We make more jokes out of it" or "We can laugh whenever we repeat the funny expressions together even after months*."

In addition, 12% of the whole videos were preferred to be watched with the associated friend category, documented by nine participants. They mostly preferred to watch documentaries with them. We found several reasons documented in diaries: seven participants reported that they would like to watch and communicate about the topic of the documentary with whomever they have common interest with. Participants also added that through this way, they can talk to someone who already knows about this topic (7P) and they may provide additional information (5P): *"It was a documentary about dancers and I was excited if I could watch it with two of*

Genre		News	Film	Soap	Comedy Series	Quiz	Documentary	Sport
Relationship Categories	Germany	Family	Nobody	Fun Friends		Associate friends		All
	South Korea		Family			Useful Contacts		

Figure 4.9: Social pattern preferences for different genres.

my friends. We are participating in the same dance class weekly." "*The documentary reported mainly about car racing. I'd like to watch it with associate friends as we professionally play online racing game" or "I'd like to watch it with her (associate friend) because we discussed about the same topic but I could not convince her about what is clearly shown in this program"*.

By analyzing genres individually, we found that the sports genre has a very similar number of videos over almost all types of people's relationships. Five participants even explicitly wished to watch sports with strangers. They believed that *"Sports means: more people, more fun"* or *"watching football in group is more enjoyable"*.

We found that only news is the most frequent genre (over 70% of documented news videos) which German participants wished to watch with their family. In suppor of this, seven participants reported that sharing and talking about news as the main reasons: *"I always inform my wife about weather conditions.*"

To help readers, a summary of preferred social connectedness for different genres of each country is given in table 4.9.

South Korea

Current Patterns of Social Watching: The Korean participants documented that they watched videos with others in 114 of videos (47% of videos). Similar to Germany, family categories have the highest frequencies for social watching in Korea. This fact appeared approximately in 56% of the samples and was mentioned by all participants. They rarely reported about their social watching experiences with weak relationships. Useful contact was the only category documented by five participants that viewed videos. They stated, *"We [colleagues] are together every day, sometimes even on weekends!"* or *"I cannot watch videos alone in my office, my colleagues are always with me there"* as reasons for watching videos with their useful contacts.

Preferences of Social Connectedness: The Korean participants preferred watching videos in 149 of videos (62% of videos). In general, they preferred to watch video with their strong relationships, such as close and far family members and, particularly confidant friends (72% of videos) while watching news (8P) or films (6P). This behavior is similar to the current collocated social watching practices discussed above. They reported *"I always talk about interesting news with my husband"*, *"The news was related to our daughter education"* or *"Movies are too long and I would prefer to share my time with my family.*"

Our analysis for the weaker relationships revealed only associated friends as the most preferred relationship for watching quiz videos (52% quiz video were indicated to watch with associated friends by seven participants). They mentioned the joy of guessing the answers (5P), presenting their knowledge (7P), and discussing about the topics of questions in quiz (6P) as the main reasons.

Nine participants documented they would like to watch sports with all types of categories. However, unlike in Germany, we did not code any preferences to watch sport with strangers. *"It would be great if I could have all of my friends (confidant, comforter, fun, and associate) present when we won the gold medal"* or *"Our national favorite player made a goal and it would be more exciting while watching it in group.*"

Patterns of Non-Social Situations

We also analyzed the videos of the category "nobody" in which the participants wished to watch alone with respect to different genres. This allowed us to gain some insights into videos and genres that were preferred to watch in non-social situations in both countries (G: 78, K: 89 videos). The results of our analysis turned out to be very similar across both countries.

We found that news (G: 7P, K: 8P) and films (G: 9P, K: 6P) are the genres, which people mainly preferred to watch alone. As we detailed above, we also coded almost the same amount of videos for preferences to watch these two genres within families (i.e., social situations). To explain this opposing findings, we further analyzed participants' comments in diaries. We found that preferring to watch films and news genres alone was mainly found in situations where they need more continuous attention (G: 5P, K: 3P) [Geerts 2008a]. *"It was a complex movie; it was hard to follow if you talk to someone.*" Other reasons for non-social watching preferences were the mood of participants (G: 10P and K: 6P) and quality of the film or news (G: 3P, K: 3P). These reasons reduced people's willingness to watch these genres together or communicating about them while watching. *"I'm in bed and it just helps me to sleep.*"

In contrast, quiz (G: 7P, K: 8P), comedy series (G: 8P, K: 10P), and sports

(G: 11P, K: 10P) are three genres, which participants stated in both countries that they would have more fun when they are watched with others or in a social situation. This can be explained by the attempt of viewers to predict the outcomes of the quiz while they watch or by the creation of ongoing discussions around comedy series to have more fun together. The participant's desires of watching sports in social situations were confirmed not only by our empirical results but also by several other researchers (e.g., by Ducheneaut et. al [Ducheneaut 2008]). Moreover, the results of non-social preferences reflected the findings of the study by Geerts et al. [Geerts 2008a] in which they investigates the role of program genres in terms of different activities, such as talking and sharing.

Limitations

While most of our findings represent generalized patterns, because of the limited number of participants, we were not able to identify possible cross-cultural differences. Moreover, our analysis is limited to what the participants actually documented in their diaries. It would be interesting to conduct cross-cultural exploration to have adequate understanding of the role of social structure on social communication between remote households and observe if there is any behavior differences in this topic from country to country.

4.2.2 Design Requirements

Based on findings of the preliminary user study and results of literature analysis reported in 3.1, we compiled four requirements as the rationale for the design of our system (CoStream@Home) supporting remote communication between households and the field.

R1. Support social pattern preferences based on genres

Based on the study results, we identified several generic patterns about the preferences of social connectedness to support the remote social watching experience (cf. figure 4.9). Our results showed that the nature of some friendships can be matched to certain genre characteristics. Different friendship categories demonstrated remarkably specific behavior related to the social watching experience. For example, associate friends are frequently preferred as coviewers despite their weak relationships. Since the nature of this category is mainly based on entirely sharing a particular interest or activity, it was highly desired to communicate while watching, especially in quiz and documentary.

As a consequence, social television systems may adapt their buddy list to the content that is being shown. To enhance the potential of communication and the fun of watching in social situations, it is necessary to design the structure of buddy list more effectively. More specifically, TV systems can better reflect the desired social structure of viewers instead of only listing accidental online friends. As another concrete example, such systems can improve their content recommendations based on the interpersonal relationships between viewers and online friends.

R2. Support activity awareness

Another interesting issue that implicitly arose in participants' comments was that they would like to follow episodic programs with the same friends, whom they watched previous episodes with. This fact shows that people desire to plan activities, manage availability, and arrange conversations with each other. This can be achieved by displaying information about the current user's activity allowing for social awareness between TV viewers whomever are not present. This awareness refer to both the presence of other noncollocated viewers and a disclosure of their activities in order to request for sharing information and establishing communication.

R3. Support more number of people from a broader social relationship spectrum while watching the sport genre

Another fact in our analysis on relationship preferences was the sport genre that had a fair distribution over all types of interpersonal relationships. It highlights the fact that for sport viewers the quantity of coviewers is more important rather than the social structure of people. This results motivated our design to focus on live sporting events – as the first step, rather than other genres (e.g., soccer matches) in order to bridge the experiential gap exists between sport viewers located at homes and in the field.

R4. Reflect the context of the user

Although we discussed that STV systems should benefit from the social structure of watching experience, our results showed a relative high desire on watching videos in non-social mode. This suggested that social iTV systems may take contextual factors, such as day time, emotion, and mood of viewers into account in order to minimize unwanted social interactions and vice versa. Therefore, it is required that the future systems will be context- and emotion-aware and utilize this information to offer TV viewers more delightful watching experience as well as desirable and

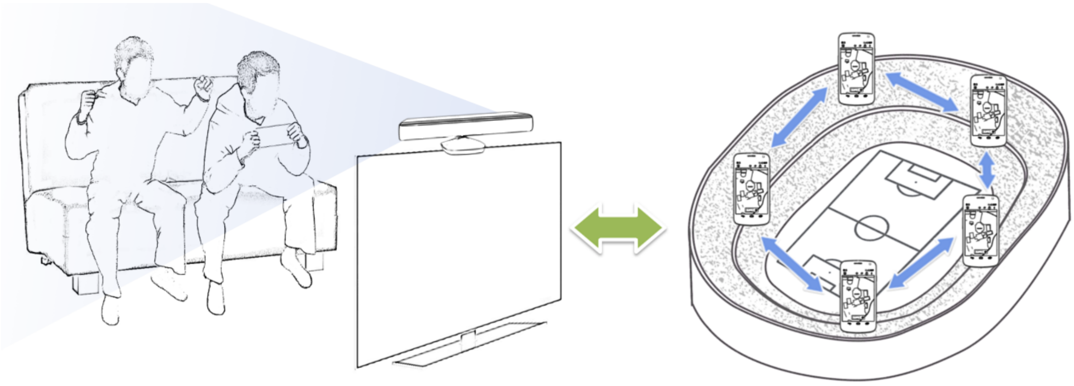


Figure 4.10: Two-way communication in real-time between the field and living rooms.

useful communication means.

Table 4.3 illustrates a summary of the requirements. Since exploring the effect of social structure preferences on all genres was not feasible in the frame of this thesis, in this chapter, we address the requirements in the context of sporting genre. Therefore, the underlying design and concept of our system (CoStream@Home) focuses on supporting live sporting events and connecting experiences between homes and the field. Nevertheless, we show later how it can support other relevant genre types as well.

4.3 CoStream@Home

In this section, we focus on bridging this experiential gap existing between people in the field and those engaging remotely during *live sporting events* for all types of interpersonal relationships (R3). To address this gap and associated challenges, we contribute a set of interaction concepts and techniques, we call *CoStream@Home*, to connect sport fans in both realms through bi-directional mobile live video sharing (cf. figure 4.10). We exploit mobile devices in both realms as a means for mutually contributing to the event engagement, potentially leading to more immersive and socially connected experiences during live sporting events.

The system particularly supports the decreased viewing perspective and facilitates social interactions and the co-construction of shared experiences across realms. In order to stimulate social interactions and eventually, enhance user experiences between spectators at home and in the field, we consider gestural information of

Requirements		Supported by prior work?	Contributions of CoStream@Home
R1	Support social structure preferences based on genres	○	CoStream@Home supports connecting the watching experience of live sport genres between homes and the event's field.
R2	Support the feeling of social presence	◐	CoStream@Home proposes a location-based real-time video broadcasting concept between users at homes and in the field. It also reflects the user's activity status.
R3	Support a broader social context while watching the sport genre	◐	CoStream@Home establishes social interactions without requiring a-prior known users and regardless of their relationship.
R4	Reflect the context of the user	◐	CoStream@Home supports implicit emotional communications based on the viewer's spatial and gestural reactions.

Table 4.3: Overview of design requirements. ○and ◐show if the state of the art have covered the requirements to some degree respectively.

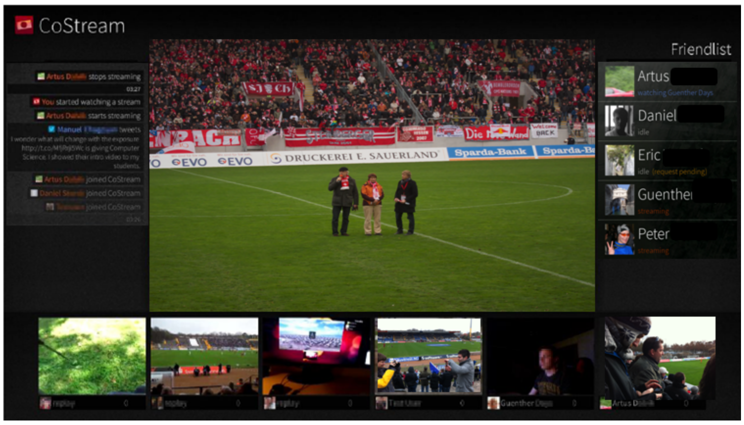


Figure 4.11: The main CoStream@Home TV application user interface.

spectators in front of the TV (e.g., emotional and gestural reactions (R4)) in addition to the live video sharing communication. We believe that such information can open up novel social interaction possibilities. The underlying system design and interaction concepts are described in the next section, followed by the implementation details of the system architecture.

4.3.1 Underlying Interface Concepts

CoStream@Home consists of three components:

(a) a main TV application running on a nearby computer connected to a Samsung Smart TV along with a Kinect camera located atop the TV,

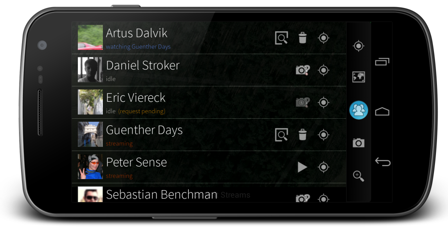
(b) an application for TV viewers at homes running on a companion device,

(c) a mobile application for users in the field

The latter is inspired by CoStream application presented in 2, supporting the connection to users at home.

Users at homes first need to log into the system on the companion device using their Facebook account. Upon a successful log in, the application connects with the TV application that is shown in figure 4.11. It displays the live professional broadcast in the main preview center, the user's buddy list, a notification list of users activities, and live broadcasting streams coming from the mobile users in the field. The users can interact with the system using the application on the companion device. We conceptually subdivided the interaction design into three modes described in the following.

(a) The location-based overview



(b) The list preview

(c) The Heat map visualization

Figure 4.12: The application running on the companion device that provides real-time awarenss about the users in the field in three different ways: (a) the location-based overview of users in the field, (b) the list preview of all users who are in the field and use the CoStream application along with their status, and (c) the heat map visualization displaying areas of interests.

4.3.1.1 Awareness and Overview across Realms

Live broadcasts are commonly restricted to professional camera perspectives. These cameras mainly cover the primary scene of the story but not the other interesting scenes. For example, reactions of bystanders or friends are not covered. Therefore, we developed a location-based real-time video broadcast between users at homes and those in the field. Mobile cameras of the users in stadiums can become “remote eyes” for viewers at homes. They can be used as cameras on-demand so that users can watch from different perspectives and be socially connected through real-time videos.

Our system initially provides an overview of the remote user on the companion device through the Google satellite map view (cf. figure 4.12 (a)). Current location and orientation of remote users in the field (who are logged into the CoStream application 2.2.2) are indicated with a custom-designed marker on the map. The marker decoration displays the Facebook profile image of the corresponding user

Figure 4.13: The TV application: three user generated video streams are visualized and the context menu is activated on the first video stream.

and reveals whether the user is idle, broadcasting, or watching a video stream on her mobile device. To get a quick overview, the user can tap on the “Users” button to open a slider containing a list of all users in the field along with their application status (cf. figure 4.12 (b)).

During time-critical sporting events, it is not easy to search and find the best user-generated viewing angel only based on the location. In order to provide a richer context (R2,R4), we provide the geographic distribution of user generated video perspectives in the form of a heat map. The color intensity in the heat map corresponds to the frequency of perspectives in the stadium. The proposed heat map can quickly and clearly show where in the stadium receive the most attraction. We designed the heat map visualization (depicted in the figure 4.12 (c)) for displaying areas of interest during live events. It can be activated by tapping on the corresponding button located on the left side of the interface. The heat map is visualized if two or more users record videos with an overlapping viewing angle.

4.3.1.2 Active Engagement and Social Interaction

The current watching practice of a live sporting event is isolated from the sport fans present in the stadium. In effect, remote users cannot contribute to the overall event experience in the stadium and vice versa. To support this, we developed a push and pull technique so that users at homes and in the field can mutually notify each other about the interesting scenes (R2). More specifically, TV viewers can request spectators to start streaming a scene from their perspective (or watching an already broadcasted stream) by tapping on their marker. The stream is then shown on the bottom of the TV screen in a picture-in-picture mode as it is shown in figure 4.13.

To interact (play/pause, relocate, zoom, or close) with the video streams shown on the TV screen, we developed a remote controlling (RC) mode on the companion device. This mode is activated when users hold the mobile phone similar to grasping the conventional TV remote control (one-handed in portrait mode and slightly slanted). Once the RC is activated, the mobile screen freezes and a blue frame is visualized around the video stream indicating the current selection. Users can then navigate to other video streams by performing simple swipe gesture on the mobile phone. To play/pause the stream users can simply tap on the screen of mobile phones. Pinch gesture enlarges the stream preview on the TV screen. Long press activates a context menu on the selected screen (cf. figure 4.13) offering the following functions:

Move : Users can arbitrarily move the video stream window across the TV screen. This feature is particularly helpful when the default position of streams (bottom of the TV screen) disrupts the professional TV broadcast.

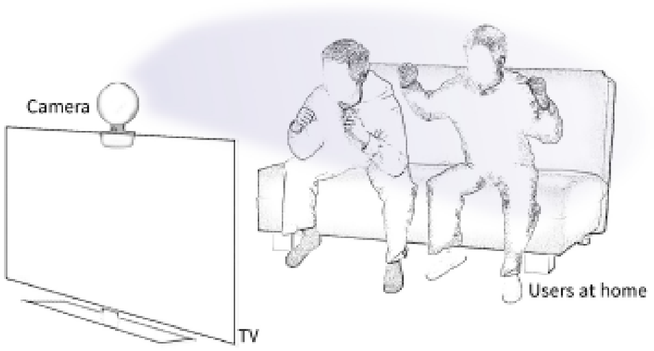
Swap : Switches the inset (the selected stream) with the full screen preview of TV (the professional broadcast). In this way, users can quickly watch the live user-generated video in full-screen mode. The professional broadcast becomes as an inset in the stream list.

Delete : Removes the stream from the TV screen.

Conversely, users in the stadium can ask TV viewers to start a video stream. Upon a stream request from the field, TV viewers receive a notification on their mobile phone and then can activate the Kinect-webcam mode to start streaming to the field. This may help them establish a bidirectional communication to enhance the watching experience.

4.3.1.3 Implicit Communications

Typical social TV systems allow viewers to explicitly share experiences through various communications channels such as text, voice or video chat [Geerts 2006b]. However, these means are time-critical and may distract viewers, for example, text chat certainly requires a lot of a user's attention [Geerts 2006b]. We believe that a concept that tries to explore the affective responses felt by remote sport viewers, while watching their favorite teams, can provide a solution to enhance the event experience at homes. We think that in order for this concept to work, it is important that the system persuades viewers to perform emotional gestures, such as clapping.



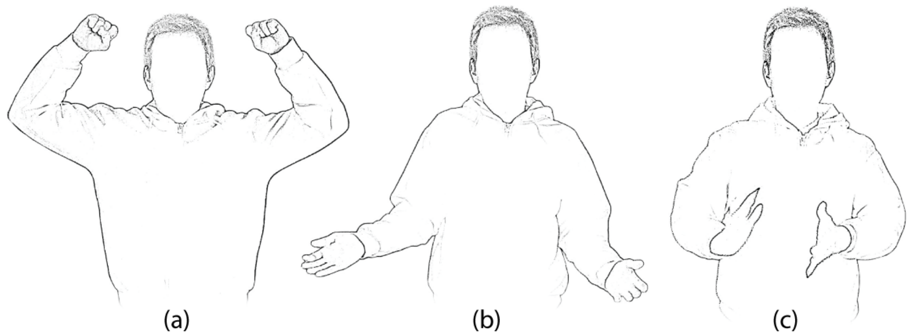


Figure 4.14: Gestures currently recognized by Kinect: (a) cheering, (b) frustration, and (c) clapping.

To do so, our system supports implicit communication means using the viewers' spatial and gestural information in front of the TV to open up novel experiences. Leveraging viewers' postures and emotions enables intuitive interactions and allows viewers to emotionally react like they do in real life without being distracted from the event.

In our system, we developed a respective interaction concept in which the reactions of TV viewers to the precious moments are implicitly transferred to sport fans in stadiums. The system continuously processes the skeleton tracking data coming from the Kinect depth camera and can recognize three common expressions:

Cheering : When users raise both hands as a sign of cheering and appreciation (cf. figure 4.14 (a)),

Frustration : When users quickly move their hands as a sign of frustration and discouraging (cf. figure 4.14 (b)),

Clapping : When users clap their hands, quickly and repeatedly to express appreciation or approval (cf. figure 4.14 (c)).

In addition to the visual notification on the mobile application, sport fans in the field are cued through vibrotactile feedback using their mobile phones in the noisy environment of the stadium. After viewers perform specific gestural action, the system automatically interprets it and transmits the emotion to the fans in the field using different vibration pattern on the mobile devices.

Since the vibration patterns are easily recognizable by their duration of the vibration and the duration of the waiting period, our system formed three distinguishable vibration patterns (e.g., clapping gesture results in a fast and continues vibration). This can potentially helps remote fans feel at home with a less degree of interruption. We argue that such real-time multi-channel communications has a great potential to enrich both social and user experiences while watching events in living rooms.

We do not claim that the above expressions are exhaustive examples of the human expressions while following a live event. We however believe that these expressions are salient and frequently performed. With the advent of more precise and high-resolution depth-sensing technology, future work should consider recognizing finer and more detailed facial expression of the TV viewers.

Table 4.4 summarizes the interaction concepts of CoStream@Home.

4.3.2 Implementation

The system is implemented in three main components: (1) a TV application, (2) an Android application for the secondary device, and (3) a centralized server. These components are depicted in figure 4.15. The TV application runs on a PC connected to a Samsung Smart TV and is implemented in Java using the JavaFX [1] framework. To render the video streams and the professional broadcast, we use the vlcj [2] framework that provides direct Java bindings to the VLC media player.

The application on the secondary device is implemented in Java for the Android platform. It automatically establishes a TCP connection to the TV application over a local wireless network. We use the Facebook API at start-up to authenticate users and synchronize their friends. Upon a successful authentication, the application sends the log-in information to both the server and the TV application.

As we discussed in Section 4.3.1, the users at home can interact with the system using the application on the companion device, which has a remote controlling

[1] http://www.oracle.com/technetwork/java/javafx/overview/index.html

[2] http://www.capricasoftware.co.uk/projects/vlcj/index.html

Name	Purpose	Description
Awareness and Overview Concepts		
Name	Purpose	Description
Location-based Overview	Support ‘remote eyes’ (Field-to-home)	Location and orientation of users are indicated with a marker on the map.
Heat-map Visualization	Display areas of interest (Field-to-home)	Overlapping viewing angle are visualized on a heat-map.
Active Engagement Concepts		
Name	Purpose	Description
Push and Pull	Support notification (Field-to-home and wise versa)	Ability to push to stream, to activate the Kinect-webcam mode and to pull to watch a stream
One-hand Remote Control	Support interaction with on-screen videos (Home feature)	controlling functions using companion device
Implicit Communications Concepts		
Name	Purpose	Description
Cheering	Support implicit chant (Home-to-field)	Notification about cheering and appreciation
Frustration	Contribute emotional expressions (Home-to-field)	Notification about frustration and discouraging
Clapping	Provide remote support to a team (Home-to-field)	Notification about appreciation or approval
Vibro-tactile Feedback	Support real-time communication (Home-to-field and wise versa)	Notification about their feeling regarding the ongoing moments.

Table 4.4: Summary of the interaction concepts and the respective interaction techniques presented in CoStream@Home.

mode. This mode is an overlay that tracks touch positions and movement of fingers including left to right, right to left, top to bottom, and bottom to top on the mobile screen. If a swipe or touch gesture is detected, the application sends the recognized

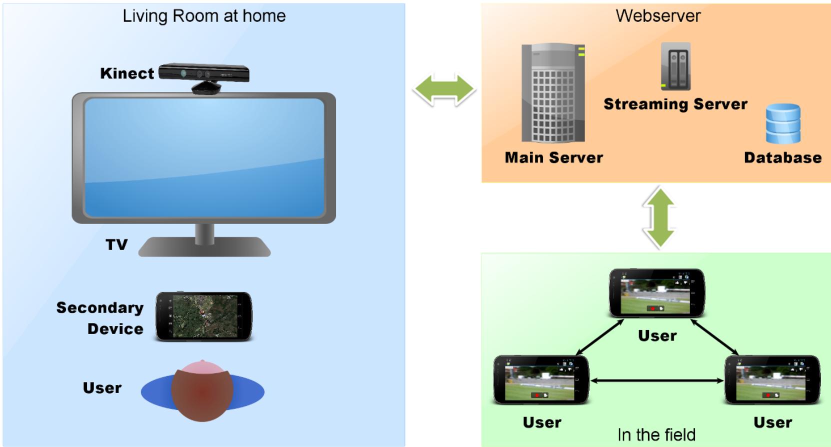


Figure 4.15: Technical Architecture of the system.

command to the TV, and consequently, the TV provides appropriate visual feedback on its screen.

The last component of the system is the centralized server. It is implemented in Java and handles the communications between homes and in-the-field clients using remote procedure calls (RPC). This has the advantage that no persistent connection is needed. Moreover, if the connection is temporary not available, the system can easily resend requests at a later time. While the server is pull-based – every information from the clients has to be pushed to the server and every information on the server has to be pulled –, it avoids issues of blocked ports and firewalls on the client side. Both the TV and the Android application in the living room are connected to the server.

Furthermore, the server is responsible for decoding incoming and encoding outgoing video streams. We run a VideoLAN Manager (VLM) [3] instance that is responsible for storing the user-generated videos on the server. Currently, we expect real time protocol (RTP)[4] streams as input and use VLM to redistribute them as HTTP streams for a better compatibility with different platforms.

4.4 Early User Feedback

We evaluated our concepts presented above in an early user feedback session. The goal of the study was to validate our concepts with professionals and obtain initial insights on the usability of the CoStream@Home user interface.

[3] http://www.videolan.org/projects/vlma/
[4] http://www.networksorcery.com/enp/protocol/rtp.htm

Study Design and Methodology

In order to get initial user feedback on the interaction concepts, we set up a lab study and recruited five experts and fellow members from the local university department (26 years of age in average, and all male). Each single-user session lasted about one hour. The study environment was designed in way that it resembles a living-room setting including a 46-inch television, a couch, and a couch table. To simulate the real-time communication between to the realms, we played a pre-recorded soccer match on the TV screen as the main TV broadcast, while the user could receive user-generated videos that were previously captured by evaluators from different viewing angles of the same event in the stadium. In addition, the participant's spatial gestures were tracked using a Microsoft Kinect positioned atop of the television. We also provide them a smartphone with the Android application of CoStream@Home installed.

In each session, we first introduced our concept briefly and then present the interaction techniques by going through a set of tasks and functionalities. This allowed participants to better understand how CoStream@Home works and gradually engage in its interaction concepts and techniques. Each task aims to cover at least one interaction techniques of the prototype so that the participant is familiar with every possible techniques. The order of the tasks was the same for each participants. A complete list of tasks is shown in table 4.5.

In the study, the participants were asked to watch the soccer match, and use the system to communicate with the other simulated persons in the event like they would normally do. The type of interaction and communication through the system were not explicitly instructed and were left up to the participants. After each session, individual interviews were conducted with each participants. Furthermore, we videotaped the sessions and asked participants to think aloud. In our analysis, we selected the quotes and videos using an iterative open, axial, and selective coding approach [Strauss 1998].

Results and Discussion

The analysis yielded four categories. We present the results within these below.

Overall User Reaction: Overall, our qualitative analysis revealed that the participants liked to use CoStream@Home while watching live sport soccer matches and were able to quickly understand the interaction techniques supported by the connected application between television and smartphone. The users stated that shar-

⊕ Start the mobile application and log into the system.
⊕ Find the user "Alice" on the map.
⊕ Locate your own position.
⊕ Use the swipe mode to start watching the stream of "Alice".
⊕ Use the swipe mode to select and highlight this stream.
⊕ Toggle the broadcast area to full screen via swipe mode.
⊕ Make some Kinect gestures (i.e., cheering, clapping, and fast-waving).
⊕ Toggle the broadcast area back to non full-screen with the slider.
⊕ Start streaming with the Kinect.
⊕ Use the friend list to remove the stream of "Alice".
⊕ Start watching the same stream that "Bob" is watching via the friend list.
⊕ Use the heatmap to display current streams at the map.
⊕ Ask "Bob" to start an own stream.
⊕ Stop own Kinect streaming.
⊕ Select and highlight all user streams.
⊕ Toggle to full screen again.
⊕ Decline and accept a streaming request.

Table 4.5: List of tasks based on each interaction techniques

ing user-generated videos, gestures, and notification messages increased the level of entertainment and excitement during the broadcasted match. As one participants put it: *"I would use this system on a real soccer match."* Another participant confirmed that *"the system makes watching the soccer match more interesting"*. They all emphasized that they like the fact that they did not have to divide their attention between different actions, such as watching the game, selecting and watching user generated streams, and having overview on friends and their activities. This was one of the most cited reasons why participants found CoStream@Home useful and viable.

Furthermore, the participants believed that the level of distraction introduced through the gestural communication in CoStream@Home is much less than traditional communication means such as, text or audio chat. The system also recognized the three emotions reliably; in 80% of cases, the system correctly classified the emotion of users. There was a strong consensuses among participants that the vibro-tactile feedback, notification of other users' status, and activities greatly helped understanding the state of application.

Design Improvements: We also received a myriad of comments and design improvements about the user interface of the TV application. The discussion with participants revealed that they were unsure about the fixed position of digital on-screen widgets (e.g., buddy list, video streams, etc.). One participant commented, *"While watching the match in full-screen mode (cf. figure 4.13), the user-generated streams sometimes occluded parts of important scenes of the match."* Another participant suggested, *"It might be more useful to visualize the information on a permanent box at the top side of the screen as the ball is usually moving on the ground at the bottom of the screen."*

We believe that this issue will become less severe by enabling users to effectively control the on-screen widgets by themselves; users can dynamically change the position, size, dimensions, and arrangements of widgets in their desired way. As another solution to this problem, we propose that users can transfer an UI widget (such as a video stream) to the companion handheld device and thus making the TV screen free of unwanted UI widgets.

Concerns and Suggestions: Almost all participants commented on the fact that they were missing interaction supports with other collocated viewers whom also use CoStream@Home. One participant added, *"I want to become aware if both me and my wife are simultaneously watching the same stream from our friend on our secondary devices. So in such situation, the system should suggest sharing it to full screen on the television to have a bigger view and open a discussion around."* We also believe that large horizontal interactive surfaces (digital tabletop computers) can be potentially an advantageous add-on to the CoStream@Home system to effectively support such collocated situations. For example, TV viewers can transfer a user-generated video stream to the tabletop and collaboratively watch and follow related messages or Twitter feeds together while having the professional broadcast displayed on the TV screen.

One participant was concerned with the small size of the screen available on the mobile devices, in contrast to other hand-held devices. While he commented: *"I prefer having the CoStream@Home application installed on a tablet PC as secondary device instead of a smart phone since it has more space for showing information"*, the other participants stated, *"Smartphones are lightweight and always available even when I leave the living room for a short time during the event. Other handheld devices such as Tablet PCs may be to exhausting to use during the whole 90 minutes soccer matches, because it is heavy and does not support the benefit of comfortable one-hand usage of smartphones."* We believe that depending on the context and user's preferences, both types of devices can be used as the companion device for

CoStream@Home.

General Findings
– CoStream@Home increased the level of entertainment and excitement during the broadcasted match.
– Gestural and emotional interactions were found as a natural way of communication among soccer fans that could minimize the distraction from the TV sport broadcast.
– There was a concern with choosing an appropriate hand-held device as the secondary screen (small-size screen of the mobile devices vs. heavy tablet PCs) for CoStream@Home.
– The CoStream@Home should support an efficient control over the on-screen widgets on the TV screen.
– The designed user interface should consider interactions within collocated viewers at homes to optimize the domestic (at home) user experience.

Table 4.6: Summary of results from the early user feedback session evaluating CoStream@Home.

4.5 Conclusion

Typically, local-scope mass events can be experienced in two different ways: spectating it on-site (e.g., in the stadium) or remotely at homes. Although both experiences concern the same event, they are fundamentally different. Spectators in the field live witness the event through both listening to the atmosphere and peripheral vision. On the other hand, people in living rooms remotely watch the event with the perspective of professional broadcasts. While they can access some additional information related to the event, they lack the live atmosphere and social interactions with the attendees in the event. In general, people at home cannot contribute expressions such as emotions to the event experience in situ and vice versa.

In the present chapter, we addressed this experiential gap between homes and the field during live events through a two-way real-time video and context sharing. We started by conducting a preliminary study to deepen our understanding about people current communication patterns in front of the TV. We especially concerned with identifying how people typically communicate with noncollocated viewers (who are either watching in another home settings or in the field). The study results helped to understand the role of interpersonal relationships on social video watching habits. We found a probable link between the nature of friendship variations and program genres that can inform the design of social television systems.

Based on the study findings, we advocated the use of user-generated mobile video sharing and leveraged viewers' gestural information to bridge the experiential gap between in situ and remote event experiences. We presented a system to address these issues, along with three interaction concepts as a first step toward supporting fundamentally new live event experiences. Therefore, we developed a location-based real-time video broadcast between users at homes and those in the field. Mobile cameras of the users in stadiums can become "remote eyes" for viewers at homes.

To support social interaction, we developed a push-and-pull mechanism so that users at homes and in the field can mutually notify each other about the interesting scenes. The third interaction concept focused on implicit communication between the two realms in which the reactions (i.e., cheering, frustration, and clapping) of TV viewers to the precious moments are implicitly transferred to sport fans in stadiums. These concepts are particularly designed to foster active engagement between spectators located in both realms, such as sharing location-based live user-generated videos between TV viewers and spectators in the field.

In early user feedback, we evaluated these techniques in the context of a live television broadcast of an event. The results show that the first steps of CoStream@Home concept was successful in bringing a selection of emotional and social experiences to fans who interact in real time from their homes without being on the same shared place. We also found that supporting the aforementioned experiences through broader kind of small handheld devices can be very useful to provide efficient and flexible control over on-screen digital graphic, and collaborative interactions for co-located fans in the same living-room while watching a soccer event.

Chapter 5

Conclusions

In this final chapter, we revisit and summarize the main contribution of each research direction and last, identify directions of future research.

5.1 Summary

Local-scope mass events are prime social and media-intensive phenomena, attracting not only a large number of spectators to directly witness and experience the life atmosphere in-situ but also an even larger number of remote viewers who follow the very specific topic through media coverages at homes. While both spectators in-situ and viewers in living-rooms concern the same event, their experiences and challenges each face to are fundamentally different. Decreased viewing perspective and social experiences imposed by the physical restriction of live events (particularly in stadiums and arenas where spectators are assigned to seats) are the main challenges that may diminish in-situ experiences.

At homes, user input to the television systems is mainly device-based and mostly require lots of viewer's visual attention for operation. This may mar living room experiences particularly when following a live event program that the viewer's attention is entirely directed to the TV screen. In addition, remote TV viewers lack the live atmosphere of events at homes and therefore, are willing to connect and communicate with event followers located in the field. However, inadequate design of interfaces and communication forms can significantly diminish the living room as well as in-situ user experience and turn them into distracting and impractical than helpful.

In response to these challenges, the goal of this thesis is to develop novel interaction concepts to support interactions, social connectivity and providing truly natural and immersive shared experiences for both spectators in the field and viewers at homes. To this end, the present thesis followed three research directions (i.e., ...) that each are based on an empirically-inductive approach to develop novel user experiences and interaction designs. In the following, we briefly recall the main outcomes and contributions of each research direction.

5.1.1 In-situ Experiences

In this direction, we addressed challenges stemming from the physical confinement of spectators in the event's venue where they are assigned to a particular seats or can not easily move from their location and thus, are limited to a certain viewing angle. This impedes social interactions and obtaining a real-time overview and awareness about the event, and activities of fellow spectators. More precisely, in this direction we aimed to enhance in-situ user experiences through the design of a novel location-aware video sharing concept so that the mobile cameras of spectators can become remote eyes of others. This allows for observing an event from various perspectives and angles.

To do so, we followed a user-centered design approach in which we first empirically and iteratively investigated how can live sharing of user-generated videos support the co-construction of experiences during events. We also analyzed and established the interface requirements for such systems through three iterative focus group sessions.

The focus group studies showed that obtaining an efficient overview and awareness of the event, enabling proper social interaction, encouraging active engagement, and providing immediate and less visually demanding interaction as four main requirements of interfaces supporting spectatorship experiences in-situ. These requirement provided a solid grounding for the design of our interface concept coined as CoStream that is evaluated in a series of field studies.

CoStream

We contributed CoStream: a set of interaction concepts that are coherently implemented as a mobile application to address the in-situ challenges of events' spectators. These concepts are particularly designed to support obtaining overview and in-situ awareness, watching and streaming, and active engagement. Based on how users naturally hold the device (parallel to the ground or upright), the user's current location and of nearby spectators is shown in a map or an augmented reality view. The latter provides in-situ awareness and is invoked when the device is lifted and held facing the environment (upright) like a see-through display.

Moreover, CoStream fosters immediate interactions as users are able to just rotate the device to start watching or broadcasting a stream. Through a pull-and-push technique, it also allows users to actively draw their friends' attention to what they are doing. Tapping onto the current video and dragging it to a friend (push mode), the friend is invited to watch the same stream as the user. And tapping onto a friend's icon and dragging it into the video screen (pull mode), the user switches to the same video the friend is currently watching or streaming.

Field Studies
The CoStream concepts and interface were evaluated in two field studies during two mid-scale soccer matches. Results showed that our concept supports the in-situ co-construction of shared experiences in three different ways. First, it enriches social and spatial awareness by enabling users to built a cognitive map of the event's location with their friends being landmarks and therefore serving as quick access shortcuts to different perspectives. Second, it encourages active spectatorship by focusing on not only consuming the other user's videos but also mainly on producing their own videos.

In addition, the users frequently pointed their friends' attention to interesting streams they were either watching or recording. Third, it enriches social interactions and feeling of connectedness between friends or even with the whole audience of their stream. In addition to these benefits, the findings also revealed a tension between the conventional physical experience of the event and the CoStream-based digital experience of the event. The results highlighted that as our concept contributes to the event through a strong real-time coupling between physical and digital experiences, this tension can be characterized as an interplay of both experiences. As a consequence, spectators can freely choose either to become more connected to other participants' perspectives through CoStream, or to the the physical atmosphere and experience of events or intertwines both experiences.

Based on findings, we concluded that the CoStream enhances and intertwines with the overall event by enabling a fluid transition between focus and context of live events through (a) providing efficient overview and awareness, supporting (b) active engagement, (c) immediate interaction and (d) reducing visual attention. In summary, the novel digital experience with CoStream 'competes' with the real-world experience.

5.1.2 At-home Experiences

In this direction, we targeted the watching experiences of people that follow live events from their homes remotely in a lean-back and much more relaxing way, compared to the in-situ experiences. We argued that while the current way of interacting with TVs (i.e. through either button- or touchscreen-based remote controls) are well-established interaction paradigm, being device-based and requiring visual attention are two main drawbacks of them particularly for time-critical interaction while watching live programs. Therefore, in this research direction, we proposed a novel interaction style between television and viewers at homes that is based on the human body due to its various advantages, such as being omnipresent, deviceless, and eyes-free. More precisely, we investigated how user input for interactive televisions can be redesigned to become more usable and offer more delightful user experience.

In order to better understand the use and movement patterns of the body in front of the TV, we first observed how people naturally watch and interact with TVs in a field study. The study findings showed that the whole body information – such as pose and orientation – have a potential to support coarse-grained TV interactions. The fine-grained TV interaction (e.g., channel navigation and selection of items) can benefit from spatial movement specified in particular by user's hand. These results provided empirical foundations for the design of two novel body based TV user interfaces – namely CouchTV and PalmRC – as the main contributions of this research direction.

CouchTV

In CouchTV, we support course-grained interactions with TV systems that rely entirely on the spatial and postural information of viewers. It contributes novel interaction techniques for (re)engaging in TV watching activity, providing appropriate level of awareness, and displaying supplementary information related to live event program. The CoachTV interface was evaluated in an initial user feedback session with twelve groups of TV viewers. Overall feedback to our concept was positive as almost all participants confirmed that they enjoyed experiencing our concept and could imagine to use it in their everyday life.

The results of the Attrakdif test showed that our concept can significantly enhance hedonic quality by 95% certainty. We also found that CouchTV fitted as an ambient display at home, supported effortless information gathering while watching and social and content connectedness, enriched interactivity with TV systems and more importantly required to be sync with on-screen content. Initial user feedback have shown that our system was appreciated and can ease (re)engagement into watching activity and coarse-grained implicit TV interactions.

PalmRC

In PalmRC Concept, we appropriated the palm of the hand as a means to enable fine-grained interactions with the TV. It is a novel eyes-free input style for television systems allowing TV viewers to perform spatial interactions with empty hands. The underlying concept of PalmRC is inspired by the sense of proprioception that enables human to sense the relative position of their limbs in the space. Unlike typical device-based TV input modalities viewers can operate television through touching the palm of their hands with the other hand index finger in an eyes-free manner. This also allows them to map remote controls functionalities to their hand and perform fine-grained interactions – such as navigation in menu using arrow keys that are mapped on appropriate salient regions of the palm.

User Studies

The PalmRC interface concepts are evaluated through a series of user studies focusing on the effectiveness and user experience of this novel TV input modality.

Exploratory study

We initially conducted an exploratory study with ten participants to empirically ground the requirements for designing an eyes-free, palm-based TV remote control. The results of the studies show that users preferred to transfer typical remote control functionalities such as directional keys to the palm (inner side) of their non-dominant hands. We also found out that the palm offers nine salient regions (landmarks), which can be easily recognized and touched without requiring any visual attention. Moreover, they preferred 2D touch gestures such as swiping on the palm surface for efficient browsing of lists with so many options. Our findings also revealed that users utilized the palm surface as a canvas to draw short symbols, such as digits or emoticons.

Controlled Experiment

Based on the results of the exploratory study, in this experiment, we verified how precisely users can touch their palm's salient regions (landmarks) without looking at them and how effectively they can select the target element of transferred on-screen user interface elements on their palm in a controlled experiment. We showed that the landmarks can be touched precisely enough for TV interaction if the size of targets is considered sufficiently large about 28mm (SD= 0.85) in diameter on the palm surface to encompass 90% of all touches. Users could reliably and effectively (>90%) map one-dimensional (1D) grid-layout menus with up to 4 options to their palm surface, independent of whether the menu is horizontally or vertically aligned.

Comparative Study

In this study, we focused on identifying respective advantages and disadvantages of PalmRC compared to the traditional remote controls and touch-based smart phones in a controlled laboratory setting. The results of this study revealed that the

Smartphone interface caused significantly more temporal demand and frustration than the other two interfaces. On the other hand, the standard remote control condition resulted in significantly less physical effort, comparing to other conditions. We believe that this is because of the two-handed nature of PalmRC (and partially Smartphone) that may require more coordination of both hands. PalmRC provided shortcuts and immediate interaction, enabled users to preserve their attention to the television and found to be practical in situations where grabbing a mediator device is difficult.

In summary, the findings of the three studies provided a fundamental basis for the concept of imaginary hand-based remote controls. Overall, we found that PalmRC provides a usable and foremost joyful way for TV remote interaction. It advances prior approaches by stimulating both pragmatic and hedonic qualities. Our observations suggest that is mainly due to its touch-based, eyes-free input characteristic, as well as the natural haptic feedback provided through one's own body parts. It is important to note that PalmRC is not meant as an alternative, but a complementary input technique for TV remote interaction.

5.1.3 Home-field (connected) Experiences

In the third research direction, we investigated how the experiential distance between TV viewers at homes and spectators in the field can be effectively bridged. Spectators, who live witness the atmosphere and peripheral vision of the event, perceive the event differently than those people in living rooms who follow the event through professional broadcasts. We started with a diary study to see how people would like to be connected and communicate with remotely located event followers. Based on its results, we contributed CoStream@Home.

Diary study

As a part of the user-centered design process, we first conducted a preliminary study in which we examine social patterns and preferences of TV viewers for remote watching with other viewers while following not only live coverage of mass events (such as sport matches) but also all other main TV genres. Based on findings, we compiled three key requirements namely: supporting social structure preferences based on genres, reflecting feeling of social presence, and supporting a broader social context while watching the sport genre. These requirements served as rationale for the design of a novel concept, called CoStream@Home to support bi-directional remote communication between households and the field during live events.

CoStream@Home

We contributed CoStream@Home: a set of interaction concepts and techniques to connect sport fans in both realms through bi-directional mobile live video sharing and gestural information of viewers in front of the TV as means for mutually contributing to the event engagement, potentially leading to more immersive and socially connected experiences during live sporting events. We also developed a respective interaction concept in which the reactions of TV viewers to the precious moments including cheering, frustration and clapping are implicitly transferred to sport fans in stadiums. In addition, we provide not only an overview of the remote user on the companion device through the Google satellite map view but also the geographic distribution of user-generated video perspectives in the form of a heat-map to support a richer context as it is not easy to search and find the best user-generated viewing angel only based on the location during a time-critical events such as sport matches.

The CoStream@Home interface was qualitatively evaluated in an early user feedback session with five HCI experts. The results showed that the level of distraction introduced through the gestural communication in CoStream@Home is much less than traditional communication means such as, text or audio chat. The Vibro-tactile feedback, notification of other users status, and activities were reported as practical features of the system, helped understanding the state of application. Moreover, participants mentioned that large horizontal interactive surfaces can be potentially an advantageous add-on to the CoStream@Home system to effectively supports collaborative interactions with other collocated viewers whom also use CoStream@Home.

5.2 Outlook and directions of Future Research

In the following, we propose three promising directions of research for further developing the concepts created in this thesis and the insights obtained as future work.

Supporting wider spectrum of spectatorship

In this thesis, we explored and supported two common ways of experiencing an event: in-situ (at stadiums, arenas, amphitheaters, etc.) and at-home in front of television sets through media coverages. But not all fans follow the event in either of these two ways. Since the 2006 Germany world cup, *Public Viewing* [1] is becoming popular and attracting huge crowds of fans to public spaces to collocatedly watch the event on large screens. This social phenomena has become a success particularly when it comes to attract crowds to watch sporting events. For example, during the

[1] http://edition.cnn.com/2010/SPORT/football/04/27/football.world.cup.germany/

2010 South Africa world cup, more than six millions people attend public viewing events in six large cities around the world [2].

Given this ever growing interest in public viewing, there is an apparent need for participants to exchange their opinions, communicate with other event followers, and express themselves as a part of a sharing community in a given time and place. While some of the concepts developed in the frame of this thesis can be adopted to support public viewing (e.g., using CoStream for in-situ communication among participants of public viewing), future research needs to be carried out to study behaviors, practices, and requirements of people attending public viewing. Moreover, evaluation of concepts and systems with real users is needed to fully grasp and understand the difference between these settings (at-homes, in-situ, and in-public) and concern appropriate technologies, interfaces, services, and interaction strategies for each in real-world.

Designing Proprioception-enhanced Input

In this thesis, we proposed leveraging the human body for operating television in an eyes-free and device-less modality. We supported coarse-grained (CoachTV) as well as fine-grained (PalmRC) interactions with TV to provide easy (re)engagement in watching activity and enable shortcuts and immediate selection of items. We outline promising research directions for both types of interaction.

With respect to the coarse-grained TV interaction, we studied the use of spatial and postural information of the whole body in single person situations. However, watching TV does not necessarily have to be a solitary experience. One area of future research needs to concern the scalability of the CoachTV concept, where many people and devices of different types may enter and leave the watching environment. Future research shall focus on how systems should react to the spatial information it gathers to create meaningful behaviors and support repairing mistakes. Moreover, given the Edward Hall's theory [Hall 1990], there are many factors such as gender, age, culture, and work hierarchies that cause different perception and interpretation of people's spatial and postural situations which needs to be considered for the design of a shared common model for whole body TV interactions.

In-terms of the fine-grained interaction, we focused on leveraging the palm surface to operate the TV, while preserving viewer's visual attention to it. This was achieved based on the sense of proprioception that enabled people to touch salient regions of their hand, precisely enough to perform basic gestures. While practical,

[2] http://www.fifa.com/aboutfifa/worldcup/

this input modality require both hands for interaction. Therefore, one promising direction for future research can focus on possible one-handed input modalities, for instance by leveraging proximity of the hand to the body [Chen 2014]. We speculate that such one-handed input modalities can potentially improve the user experience and increasing the interaction space for TV viewers.

As one concrete solution, we propose leveraging the two degrees of freedom offered by the elbow joint – i.e., flexion and extension–, to be leveraged for a proximity-based hand input interaction in the space in front of the user [Mueller 2015b]. This can be used for interaction with a linear layer-based information space alongside the user's line of sight. In this way, users can move their hands towards or away from his head to browse through successive layers presented on the TV screen.

Integrating other application domains

Parts of this thesis focused on location-based user-generated video sharing in real-time during live events. The concepts created in this direction of research and the insights obtained can be used and tailored for a variety of other application domains than covering a live event proceedings. One potential application domain is surveillance and security systems that have received considerable attention in recent decades. Basically, such systems consist of a number of fixed-positioned CCTV or IP cameras connected via cable or wireless network to a central system where all video streams are displayed. Security personnel monitor the space and surrounding using these fixed-positioned cameras. Despite their advantages, it is rarely possible to cover the complete space because 1) is not cost-effective and, 2) it violets the privacy of citizens. User-generated videos captured by ordinary people located on-site can greatly help covering an incident from different perspectives. Therefore, a promising direction of future work should study how mobile user-generated videos can be effectively integrated in surveillance systems.

In a research project called *CoSecure* we started to explore this topic. The project aims at leveraging user-generated mobile videos streamed by citizens – i.e., ordinary people – who want to voluntary help as surveillance cameras to better and more flexible monitor the space and cover an incident from various perspectives [Mueller 2015a]. In the following we briefly describe the idea and initial activities of this project to better highlight future research in this direction.

CoSecure

CoSecure is an on-going (at the time of this writing) research project in which we investigate how mobile video streams from the ordinary people in the scene can help covering of and responding to an incident. More precisely, we envision two key scenarios to be addressed in this project:

1 Support for the security centers: where all the camera views are monitored by authorities. This part of the project focuses on developing novel interface and visualization techniques for the systems where are the information from the field are gathered and visualized. Furthermore, fusion of supplementary information resources (such as data coming from various sensors types) with the video streams will be investigated.

2 Support for the mobile units in-the-field: that live capture the scene using mobile phones. In this part of the project cross-platform applications for mobile phones will be conceptualized and developed to be used by either ordinary people or authorities in the field.

Both scenarios are the main focus of the project since in the case of an incident, it is very crucial to efficiently get a proper overview of the situation. With respect to the first scenario, the goal is to see how to visualize all camera views (be fixed positioned cameras or be mobile phones coming from the field) in an interface in a location-aware fashion. With respect to the second scenario, our main objective is to design and develop mobile interaction concepts so that ordinary citizens can stream video as well as send additional information (such as tagging of the video content) to the center.

As the first step, we have defined a set of design requirements that are derived from a user study with professionals in the field of surveillance systems. Based on the requirements, we designed and developed three system applications for:

- the security center, where all the camera views – such as surveillance camera streams and mobile video streams – are monitored by authorities. We visualized all the information resources onto a 3D model substrate on an interactive tabletop allowing for arbitrary viewing,

- mobile security units in-the-field with a cross-platform mobile application that officers could interact with it to get a quick and extensive overview of the situation when the situation is critical and,

- seamlessly integrating location-aware user generated content on ordinary people's mobile phones.

The concept and the system prototype were evaluated concerning localization and interaction techniques with an expert group. The evaluation showed that usage of mobile user-generated live videos can ease and accelerate the progress of transmitting necessary information in an emergency case to higher instances. We believe that a long-term field study is needed to fully grasp and understand the usefulness of the presented applications for surveillance and security as well as other promising application domains.

Bibliography

[Annett 2011] Michelle Annett, Tovi Grossman, Daniel Wigdor and George Fitzmaurice. *Medusa: A Proximity-aware Multi-touch Tabletop.* In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11, pages 337–346, New York, NY, USA, 2011. ACM. (Cited on page 61.)

[Ault 2008] Aaron Ault, James V. Krogmeier, Steven R. Dunlop and Edward J. Coyle. *eStadium: The Mobile Wireless Football Experience.* In Abdelhamid Mellouk, Jun Bi, Guadalupe Ortiz, Dickson K. W. Chiu and Manuela Popescu, editeurs, ICIW, pages 644–649. IEEE Computer Society, 2008. (Cited on page 21.)

[Ballendat 2010] Till Ballendat, Nicolai Marquardt and Saul Greenberg. *Proxemic Interaction: Designing for a Proximity and Orientation-aware Environment.* In ACM International Conference on Interactive Tabletops and Surfaces, ITS '10, pages 121–130, New York, NY, USA, 2010. ACM. (Cited on page 61.)

[Barkhuus 2008] Louise Barkhuus and Tobias Jørgensen. *Engaging the Crowd: Studies of Audience-performer Interaction.* In CHI '08 Extended Abstracts on Human Factors in Computing Systems, CHI EA '08, pages 2925–2930, New York, NY, USA, 2008. ACM. (Cited on page 24.)

[Baudisch 2001] Patrick Baudisch, Nathaniel Good and Paul Stewart. *Focus Plus Context Screens: Combining Display Technology with Visualization Techniques.* In Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, UIST '01, pages 31–40, New York, NY, USA, 2001. ACM. (Cited on page 53.)

[Bentley 2008] Frank R. Bentley and Crysta J. Metcalf. *Location and Activity Sharing in Everyday Mobile Communication.* In CHI '08 Extended Abstracts on Human Factors in Computing Systems, CHI EA '08, pages 2453–2462, New York, NY, USA, 2008. ACM. (Cited on page 21.)

[Bentley 2009] Frank R. Bentley and Michael Groble. *TuVista: Meeting the Multimedia Needs of Mobile Sports Fans.* In Proceedings of the 17th ACM International Conference on Multimedia, MM '09, pages 471–480, New York, NY, USA, 2009. ACM. (Cited on pages 21 and 24.)

[Bernhaupt 2008] Regina Bernhaupt, Marianna Obrist, Astrid Weiss, Elke Beck and Manfred Tscheligi. *Trends in the Living Room and Beyond: Results from Ethnographic Studies Using Creative and Playful Probing.* Comput. Entertain., vol. 6, no. 1, pages 5:1–5:23, May 2008. (Cited on pages 84, 99 and 123.)

[Bernstein 2010] Michael S. Bernstein, Bongwon Suh, Lichan Hong, Jilin Chen, Sanjay Kairam and Ed H. Chi. *Eddi: Interactive Topic-based Browsing of Social Status Streams.* In Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology, UIST '10, pages 303–312, New York, NY, USA, 2010. ACM. (Cited on page 19.)

[Brutti 2010] Alessio Brutti, Luca Cristoforetti, Walter Kellermann, Lutz Marquardt and Maurizio Omologo. *WOZ acoustic data collection for interactive TV.* Language Resources and Evaluation, vol. 44, no. 3, pages 205–219, 2010. (Cited on page 64.)

[Bryant 2000] Jennings Bryant and Arthur A. Raney. *Sports on the screen.* In Media entertainment: The psychology of its appeal. LEA's communication series., 2000. (Cited on page 118.)

[Bulterman 2013] Dick C. A. Bulterman, Pablo Cesar and Rodrigo Laiola Guimarães. *Socially-aware Multimedia Authoring: Past, Present, and Future.* ACM Trans. Multimedia Comput. Commun. Appl., vol. 9, no. 1s, pages 35:1–35:23, October 2013. (Cited on page 26.)

[Centieiro 2013] Pedro Centieiro. *Bringing the sport stadium atmosphere to remote fans.* In Alejandro Jaimes, Nicu Sebe, Nozha Boujemaa, Daniel Gatica-Perez, David A. Shamma, Marcel Worring and Roger Zimmermann, editeurs, ACM Multimedia, pages 1063–1066. ACM, 2013. (Cited on pages 122 and 123.)

[Cesar 2009] Pablo Cesar, Dick C. A. Bulterman, Jack Jansen, David Geerts, Hendrik Knoche and William Seager. *Fragment, Tag, Enrich, and Send: Enhancing Social Sharing of Video.* ACM Trans. Multimedia Comput. Commun. Appl., vol. 5, no. 3, pages 19:1–19:27, August 2009. (Cited on page 71.)

[Chakrabarti 2011] Deepayan Chakrabarti and Kunal Punera. *Event Summarization using Tweets*, 2011. (Cited on page 20.)

[Chen 2010] Ming-yu Chen, Lily Mummert, Padmanabhan Pillai, Alexander Hauptmann and Rahul Sukthankar. *Controlling Your TV with Gestures.* In Proceedings of the International Conference on Multimedia Information Retrieval, MIR '10, pages 405–408, New York, NY, USA, 2010. ACM. (Cited on page 64.)

[Chen 2014] Xiang 'Anthony' Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff and Scott Hudson. *Around-body Interaction: Sensing & Interaction Techniques for Proprioception-enhanced Input with Mobile Devices.* In Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14, pages 287–290, New York, NY, USA, 2014. ACM. (Cited on page 159.)

[Chorianopoulos 2004] Konstantinos Chorianopoulos and Diomidis Spinellis. *Affective usability evaluation for an interactive music television channel.* Computers in Entertainment (CIE), vol. 2, no. 3, pages 14–14, 2004. (Cited on page 63.)

[Chuah 2003] Mei Chuah. *Reality Instant Messaging: Injecting a Dose of Reality into Online Chat.* In CHI '03 Extended Abstracts on Human Factors in Computing Systems, CHI EA '03, pages 926–927, New York, NY, USA, 2003. ACM. (Cited on page 26.)

[Coppens 2004] Toon Coppens, Lieven Trappeniers and Marc Godon. *AmigoTV: Towards a social TV experience.* 2004. (Cited on page 123.)

[Counts 2004] Scott Counts and Eric Fellheimer. *Supporting Social Presence Through Lightweight Photo Sharing on and off the Desktop.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04, pages 599–606, New York, NY, USA, 2004. ACM. (Cited on page 25.)

[De Silva 1997] Liyanage De Silva, T. Miyasato and Robbie Nakatsu. *Facial emotion recognition using multi-modal information.* In Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on, volume 1, pages 397–401 vol.1, Sep 1997. (Cited on page 120.)

[Dezfuli 013] Niloofar Dezfuli, Jochen Huber, Khalilbeigi Mohammadreza and Max Mühlhäuser. *Towards Connected Live Event Experiences.* In In Proc. Exploring and Enhancing the UX for TV Workshop at CHI, 2013. (Cited on page 115.)

[Dezfuli 2011] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Max Mühlhäuser and David Geerts. *A study on interpersonal relationships for social interactive* *television*. In 9th International Interactive TV&Video Conference, EuroITV '11, Lisbon, Portugal, June 29 - July 01, 2011, pages 21–24, 2011. (Cited on pages 12 and 115.)

[Dezfuli 2012a] Niloofar Dezfuli, Jochen Huber, Simon Olberding and Max Mühlhäuser. *CoStream: In-situ Co-construction of Shared Experiences Through Mobile Video Sharing During Live Events*. In CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12, pages 2477–2482, New York, NY, USA, 2012. ACM. (Cited on pages 12 and 17.)

[Dezfuli 2012b] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller and Max Mühlhäuser. *Leveraging the Palm Surface As an Eyes-free Tv Remote Control*. In CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12, pages 2483–2488, New York, NY, USA, 2012. ACM. (Cited on pages 12 and 60.)

[Dezfuli 2012c] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller and Max Mühlhäuser. *PalmRC: Imaginary Palm-based Remote Control for Eyes-free Television Interaction*. In Proceedings of the 10th European Conference on Interactive Tv and Video, EuroiTV '12, pages 27–34, New York, NY, USA, 2012. ACM. (Cited on pages 12 and 60.)

[Dezfuli 2012d] Niloofar Dezfuli, Manolis Pavlakis, Florian Müller, Mohammadreza Khalibeigi and Max Mühlhäuser. *CouchTV: Leveraging the Spatial Information of Viewers for Social Interactive Television Systems - Poster Presentation*. In 10th European Conference on Interactive TV (EuroiTV'12), July 2012. (Cited on pages 12 and 60.)

[Dezfuli 2013a] Niloofar Dezfuli, Sebastian Günther, Mohammadreza Khalilbeigi, Max Mühlhäuser and Jochen Huber. *CoStream@Home: connected live event experiences,*. In In proc. SAM Workshop at ACM Multimedia, 2013. (Cited on pages 12 and 115.)

[Dezfuli 2013b] Niloofar Dezfuli, Jochen Huber, Mohammadreza Khalilbeigi and Max Mühlhäuser. *Towards Connected Live Event Experiences*. In Exploring and Enhancing the User Experience for TV, Workshop in conjunction with CHI13, 2013. (Cited on page 12.)

[Dezfuli 2014] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, M"urat "Ozkorkmaz and Max Mühlhäuser. *PalmRC: leveraging the palm surface as* *an imaginary eyes-free television remote control.* Behaviour and Information Technology, vol. 33, no. 8, pages 829–843, 2014. (Cited on pages 12 and 60.)

[Dezuli 2013] Niloofar Dezuli, Jochen Huber, Elizabeth F. Churchill and Max Mühlhäuser. *CoStream: Co-construction of Shared Experiences Through Mobile Live Video Sharing.* In Proceedings of the 27th International BCS Human Computer Interaction Conference, BCS-HCI '13, pages 6:1–6:10, Swinton, UK, UK, 2013. British Computer Society. (Cited on pages 12 and 17.)

[Diakopoulos 2010] Nicholas Diakopoulos, Mor Naaman and Funda Kivran-swaine. *Diamonds in the Rough: Social Media Visual Analytics for Journalistic Inquiry.* In Proceedings of Conference on Visual Analytics Science and Technology (VAST), 2010. (Cited on page 18.)

[Ducheneaut 2008] Nicolas Ducheneaut, Robert J. Moore, Lora Oehlberg, James D. Thornton and Eric Nickell. *Social TV: Designing for Distributed, Sociable Television Viewing.* International Journal of Human-Computer Interaction, vol. 24, no. 2, pages 136–154, 2008. (Cited on page 134.)

[Edwardson 2014] Krik Edwardson. *HbbTV: Reinventing the broadcast TV UX.* Whitepapers, 2014. (Cited on page 3.)

[Engström 2012] Arvid Engström, Mark Perry and Oskar Juhlin. *Amateur Vision and Recreational Orientation:: Creating Live Video Together.* In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12, pages 651–660, New York, NY, USA, 2012. ACM. (Cited on page 118.)

[Fouse 2011] Adam Fouse, Nadir Weibel, Edwin Hutchins and James D. Hollan. *ChronoViz: A System for Supporting Navigation of Time-coded Data.* In CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, pages 299–304, New York, NY, USA, 2011. ACM. (Cited on page 41.)

[Freeman 1994] William T. Freeman and Craig D. Weissman. *Television Control by Hand Gestures.* Rapport technique TR94-24, MERL - Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, December 1994. (Cited on page 64.)

[Frohlich 2002] David Frohlich, Allan Kuchinsky, Celine Pering, Abbe Don and Steven Ariss. *Requirements for Photoware.* In Proceedings of the 2002 ACM

Conference on Computer Supported Cooperative Work, CSCW '02, pages 166–175, New York, NY, USA, 2002. ACM. (Cited on page 23.)

[Gaver 1999] Bill Gaver, Tony Dunne and Elena Pacenti. *Design: Cultural Probes.* interactions, vol. 6, no. 1, pages 21–29, January 1999. (Cited on page 123.)

[Geerts 2006a] David Geerts. *Comparing Voice Chat and Text Chat in a Communication Tool for Interactive Television.* In Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles, NordiCHI '06, pages 461–464, New York, NY, USA, 2006. ACM. (Cited on page 123.)

[Geerts 2006b] David Geerts. *Comparing Voice Chat and Text Chat in a Communication Tool for Interactive Television.* In Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles, NordiCHI '06, pages 461–464, New York, NY, USA, 2006. ACM. (Cited on page 141.)

[Geerts 2008a] David Geerts, Pablo Cesar and Dick Bulterman. *The Implications of Program Genres for the Design of Social Television Systems.* In Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video, UXTV '08, pages 71–80, New York, NY, USA, 2008. ACM. (Cited on pages 63, 133 and 134.)

[Geerts 2008b] David Geerts, Pablo Cesar and Dick Bulterman. *The Implications of Program Genres for the Design of Social Television Systems.* In Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video, UXTV '08, pages 71–80, New York, NY, USA, 2008. ACM. (Cited on pages 124 and 125.)

[Geerts 2009] David Geerts and Dirk De Grooff. *Supporting the Social Uses of Television: Sociability Heuristics for Social Tv.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, pages 595–604, New York, NY, USA, 2009. ACM. (Cited on page 123.)

[Gilbert 2009] Eric Gilbert and Karrie Karahalios. *Predicting Tie Strength with Social Media.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, pages 211–220, New York, NY, USA, 2009. ACM. (Cited on page 128.)

[Granovetter 1973] Mark Granovetter. *The Strength of Weak Ties.* The American Journal of Sociology, vol. 78, no. 6, pages 1360–1380, 1973. (Cited on page 128.)

[Gustafson 2010] Sean Gustafson, Daniel Bierwirth and Patrick Baudisch. *Imaginary Interfaces: Spatial Interaction with Empty Hands and Without Visual Feedback*. In Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology, UIST '10, pages 3–12, New York, NY, USA, 2010. ACM. (Cited on page 66.)

[Gustafson 2011] Sean Gustafson, Christian Holz and Patrick Baudisch. *Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device*. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11, pages 283–292, New York, NY, USA, 2011. ACM. (Cited on pages 66, 67, 86, 99, 100, 101 and 102.)

[Hagedorn 2008] Joey Hagedorn, Joshua Hailpern and Karrie G. Karahalios. *VCode and VData: Illustrating a New Framework for Supporting the Video Annotation Workflow*. In Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '08, pages 317–321, New York, NY, USA, 2008. ACM. (Cited on page 41.)

[Hall 1990] E.T. Hall. The hidden dimension. A Doubleday anchor book. Anchor Books, 1990. (Cited on pages 62 and 158.)

[Harboe 2008] Gunnar Harboe, Crysta J. Metcalf, Frank Bentley, Joe Tullio, Noel Massey and Guy Romano. *Ambient Social Tv: Drawing People into a Shared Experience*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, pages 1–10, New York, NY, USA, 2008. ACM. (Cited on page 123.)

[Harrison 2008] Chris Harrison and Anind K. Dey. *Lean and Zoom: Proximity-aware User Interface and Content Magnification*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, pages 507–510, New York, NY, USA, 2008. ACM. (Cited on page 61.)

[Harrison 2010] Chris Harrison, Desney Tan and Dan Morris. *Skinput: Appropriating the Body As an Input Surface*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM. (Cited on pages 62, 65 and 100.)

[Harrison 2011] Chris Harrison, Hrvoje Benko and Andrew D. Wilson. *OmniTouch: Wearable Multitouch Interaction Everywhere*. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11, pages 441–450, New York, NY, USA, 2011. ACM. (Cited on page 66.)

[Harrison 2012] Chris Harrison, Shilpa Ramamurthy and Scott E. Hudson. *On-body Interaction: Armed and Dangerous*. In Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction, TEI '12, pages 69–76, New York, NY, USA, 2012. ACM. (Cited on page 66.)

[Hart 1988] Sandra Hart and Lowell Staveland. *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*. In Human mental workload, pages 139–183. P.A. Hancock and N. Meshkati (Eds.), Amsterdam: Elsevier, 1988. (Cited on page 105.)

[Hassenzahl 2003] Marc Hassenzahl, Michael Burmester and Franz Koller. *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In Gerd Szwillus and Jürgen Ziegler, editeurs, Mensch & Computer 2003: Interaktion in Bewegung, pages 187–196, Stuttgart, 2003. B. G. Teubner. (Cited on pages 79, 80 and 105.)

[Hawkins 2005] Robert P. Hawkins, Suzanne Pingree, Jacqueline Hitchon, Barry Radler, Bradley W. Gorham, Leeann Kahlor, Eileen Gilligan, Ronald C. Serlin, Toni Schmidt, Prathana Kannaovakun and Gudbjorg H. Kolbeins. *What Produces Television Attention and Attention Style? Genre, Situation, and Individual Differences as Predictors*. Human Communication Research, vol. 31, no. 1, pages 162–187, January 2005. (Cited on pages 63, 71 and 84.)

[Hobson 2008] Dorothy Hobson. Channel 4: The early years and the jeremy isaacs legacy. I. B. Tauris, 2008. (Cited on page 63.)

[Hollan 1992] Jim Hollan and Scott Stornetta. *Beyond Being There*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '92, pages 119–125, New York, NY, USA, 1992. ACM. (Cited on pages 16 and 29.)

[Holmquist 1999] LarsErik Holmquist, Jennica Falk and Joakim WigstrÃ¶m. *Supporting group collaboration with interpersonal awareness devices*. Personal Technologies, vol. 3, no. 1-2, pages 13–21, 1999. (Cited on page 20.)

[Huber 2012] Jochen Huber. *Mobile Interaction with Large Multimedia Information Spaces*. PhD thesis, Technische Universität Darmstadt, Dezember 2012. (Cited on page 10.)

[Igarashi 2001] Takeo Igarashi and John F. Hughes. *Voice As Sound: Using Nonverbal Voice Input for Interactive Control*. In Proceedings of the 14th Annual

ACM Symposium on User Interface Software and Technology, UIST '01, pages 155–156, New York, NY, USA, 2001. ACM. (Cited on page 64.)

[Jääskeläinen 2001] K. Jääskeläinen and Taideteollinen korkeakoulu. Strategic questions in the development of interactive television programs: Dissertation. Ilmari design, 2001. (Cited on page 3.)

[Jacucci 2005] Giulio Jacucci, Antti Oulasvirta, Antti Salovaara and Risto Sarvas. *Supporting the Shared Experience of Spectators Through Mobile Group Media*. In Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, GROUP '05, pages 207–216, New York, NY, USA, 2005. ACM. (Cited on page 3.)

[Jacucci 2007a] Giulio Jacucci, Antti Oulasvirta, Tommi Ilmonen, John Evans and Antti Salovaara. *Comedia: Mobile Group Media for Active Spectatorship*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, pages 1273–1282, New York, NY, USA, 2007. ACM. (Cited on page 5.)

[Jacucci 2007b] Giulio Jacucci, Antti Oulasvirta and Antti Salovaara. *Active Construction of Experience Through Mobile Media: A Field Study with Implications for Recording and Sharing*. Personal Ubiquitous Comput., vol. 11, no. 4, pages 215–234, April 2007. (Cited on pages 5, 26 and 29.)

[Ju 2008] Wendy Ju, Brian A. Lee and Scott R. Klemmer. *Range: Exploring Implicit Interaction Through Electronic Whiteboard Design*. In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08, pages 17–26, New York, NY, USA, 2008. ACM. (Cited on page 61.)

[Juhlin 2010] Oskar Juhlin, Arvid Engström and Erika Reponen. *Mobile Broadcasting: The Whats and Hows of Live Video As a Social Medium*. In Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '10, pages 35–44, New York, NY, USA, 2010. ACM. (Cited on pages 30 and 118.)

[Juhlin 2013] Oskar Juhlin and Elin Önnevall. *On the Relation of Ordinary Gestures to TV Screens: General Lessons for the Design of Collaborative Interactive Techniques*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pages 919–930, New York, NY, USA, 2013. ACM. (Cited on page 122.)

[Kaheel 2009] Ayman Kaheel, Motaz El-Saban, Mahmoud Refaat and Mostafa Ezz. *Mobicast: A System for Collaborative Event Casting Using Mobile Phones.* In Proceedings of the 8th International Conference on Mobile and Ubiquitous Multimedia, MUM '09, pages 7:1–7:8, New York, NY, USA, 2009. ACM. (Cited on pages 118 and 119.)

[Kuester 2005] Falko Kuester, Michelle Chen, Mark Phair and Carsten Mehring. *Towards Keyboard Independent Touch Typing in VR.* In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST '05, pages 86–95, New York, NY, USA, 2005. ACM. (Cited on pages 65 and 100.)

[Laiola Guimarães 2011a] Rodrigo Laiola Guimarães, Pablo Cesar, Dick C.A. Bulterman, Ian Kegel and Peter Ljungstrand. *Social Practices around Personal Videos using the Web.* In Proceedings of the ACM Web Science Conference (WebSci 2011), WebSci '11, New York, NY, USA, 2011. ACM. (Cited on page 119.)

[Laiola Guimarães 2011b] Rodrigo Laiola Guimarães, Pablo Cesar, Dick C.A. Bulterman, Vilmos Zsombori and Ian Kegel. *Creating Personalized Memories from Social Events: Community-based Support for Multi-camera Recordings of School Concerts.* In Proceedings of the 19th ACM International Conference on Multimedia, MM '11, pages 303–312, New York, NY, USA, 2011. ACM. (Cited on pages 26 and 27.)

[Lenman 2002] Sören Lenman, Lars Bretzner and Björn Thuresson. *Using marking menus to develop command sets for computer vision based hand gesture interfaces.* In Proceedings of the second Nordic conference on Human-computer interaction, pages 239–242. ACM, 2002. (Cited on page 64.)

[Li 2010] Frank Chun Yat Li, David Dearman and Khai N. Truong. *Leveraging Proprioception to Make Mobile Phones More Accessible to Users with Visual Impairments.* In Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '10, pages 187–194, New York, NY, USA, 2010. ACM. (Cited on pages 84 and 87.)

[Liu 2007] Yiming Liu, Peter Shafton, David A. Shamma and Jeannie Yang. *Zync: The Design of Synchronized Video Sharing.* In Proceedings of the 2007 Conference on Designing for User eXperiences, DUX '07, pages 12:1–12:8, New York, NY, USA, 2007. ACM. (Cited on page 26.)

[Ludvigsen 2010] Martin Ludvigsen and Rune Veerasawmy. *Designing technology for active spectator experiences at sporting events.* In Margot Brereton, Stephen Viller and Ben Kraal, editeurs, OZCHI, pages 96–103. ACM, 2010. (Cited on page 24.)

[Lull 1982] James Lull. *How families select television programs: A mass-observational study.* Journal of Broadcasting & Electronic Media, vol. 26, no. 4, pages 801–811, 1982. (Cited on page 7.)

[Mäkelä 2000] Ann Mäkelä, Verena Giller, Manfred Tscheligi and Reinhard Sefelin. *Joking, Storytelling, Artsharing, Expressing Affection: A Field Trial of How Children and Their Social Network Communicate with Digital Images in Leisure Time.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00, pages 548–555, New York, NY, USA, 2000. ACM. (Cited on page 22.)

[Mäntyjärvi 2004] Jani Mäntyjärvi, Juha Kela, Panu Korpipää and Sanna Kallio. *Enabling Fast and Effortless Customisation in Accelerometer Based Gesture Interaction.* In Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia, MUM '04, pages 25–31, New York, NY, USA, 2004. ACM. (Cited on page 64.)

[Marcus 2011] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden and Robert C. Miller. *Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pages 227–236, New York, NY, USA, 2011. ACM. (Cited on page 19.)

[Mistry 2009] Pranav Mistry and Pattie Maes. *SixthSense: A Wearable Gestural Interface.* In ACM SIGGRAPH ASIA 2009 Sketches, SIGGRAPH ASIA '09, pages 11:1–11:1, New York, NY, USA, 2009. ACM. (Cited on pages 62 and 65.)

[Morley 1986] D. Morley. Family television: Cultural power and domestic leisure. A Comedia book. Routledge, 1986. (Cited on page 63.)

[Mueller 2015a] Florian Mueller, Sebastian Guenther, Niloofar Dezfuli, Mohammadreza Khalilbeigi, Manolis Pavlakis and Max Muelhaeuser. *CoCenter: Developing an Interactive 3D Visualization for Video Monitoring Centers.* In Submitted to the 12th International Conference on Information Systems for Crisis Response and Management. ISCRAM, 2015. (Cited on page 159.)

[Mueller 2015b] Florian Mueller, Mohammadreza Khalilbeigi, Niloofar Dezfuli, Alireza Sahami Shirazi and Max Muelhaeuser. *A Study on Proximity based Hand Input for On-body User Interfaces.* In Submitted to the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2015. ACM. (Cited on page 159.)

[Nilsson 2004] Andreas Nilsson. *Using IT to Make Place in Space: Evaluating Mobile Technology Support for Sport Spectators.* In Timo Leino, Timo Saarinen and Stefan Klein, editeurs, ECIS, pages 1398–1409, 2004. (Cited on page 25.)

[Norman 1986] Donald A. Norman and Stephen W. Draper. User centered system design; new perspectives on human-computer interaction. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1986. (Cited on page 10.)

[O'Hara 2007] Kenton O'Hara, April Slayden Mitchell and Alex Vorbau. *Consuming Video on Mobile Devices.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, pages 857–866, New York, NY, USA, 2007. ACM. (Cited on page 2.)

[Peltonen 2007] Peter Peltonen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, Carmelo Ardito, Petri Saarikko and Vikram Batra. *Extending Large-scale Event Participation with User-created Mobile Media on a Public Display.* In Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia, MUM '07, pages 131–138, New York, NY, USA, 2007. ACM. (Cited on page 23.)

[Pemberton 2003] Lyn Pemberton and Richard Griffiths. *Usability Evaluation Techniques for Interactive Television.* HCI International conference'03, 2003. (Cited on page 99.)

[Picard 2000] Rosalind Picard. *Toward Computers That Recognize and Respond to User Emotion.* IBM Syst. J., vol. 39, no. 3-4, pages 705–719, July 2000. (Cited on pages 120 and 121.)

[Raney 2006] A.A. Raney and J. Bryant. Handbook of sports and media. Routledge Communication Series. Taylor & Francis, 2006. (Cited on page 117.)

[Reeves 2005] Stuart Reeves, Steve Benford, Claire O'Malley and Mike Fraser. *Designing the Spectator Experience.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05, pages 741–750, New York, NY, USA, 2005. ACM. (Cited on page 26.)

[Russell 1989] James A. Russell, Anna Weiss and Gerald A. Mendelsohn. *Affect Grid: A single-item scale of pleasure and arousal.* Journal of Personality and Social Psychology, vol. 57, no. 3, pages 493–502, sep 1989. (Cited on pages 120 and 121.)

[Sahami Shirazi 2011] Alireza Sahami Shirazi, Michael Rohs, Robert Schleicher, Sven Kratz, Alexander Müller and Albrecht Schmidt. *Real-time Nonverbal Opinion Sharing Through Mobile Phones During Sports Events.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pages 307–310, New York, NY, USA, 2011. ACM. (Cited on pages 26 and 27.)

[Sahami Shirazi 2013] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Kai Kunze and Albrecht Schmidt. *Upright or Sideways?: Analysis of Smartphone Postures in the Wild.* In Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services, MobileHCI '13, pages 362–371, New York, NY, USA, 2013. ACM. (Cited on pages 29 and 36.)

[Scott 2003] Stacey D. Scott. *Territory-Based Interaction Techniques for Tabletop Collaboration*, 2003. (Cited on page 62.)

[Shamma 2009] David A. Shamma, Lyndon Kennedy and Elizabeth F. Churchill. *Tweet the Debates: Understanding Community Annotation of Uncollected Sources.* In Proceedings of the First SIGMM Workshop on Social Media, WSM '09, pages 3–10, New York, NY, USA, 2009. ACM. (Cited on page 19.)

[Spencer 2006] Liz Spencer and Ray Pahl. Rethinking friendship: Hidden solidarities today. Princeton University Press, 2006. (Cited on pages 126, 128 and 131.)

[Strauss 1998] A. Strauss and J.M. Corbin. Basics of qualitative research: Techniques and procedures for developing grounded theory. SAGE Publications, 1998. (Cited on pages 32, 42, 68, 79, 86, 105 and 146.)

[Sun 2007] Xu Sun and Andrew May. *Mobile Personalization at Large Sports Events User Experience and Mobile Device Personalization.* In Proceedings of the 2Nd International Conference on Usability and Internationalization, UI-HCII'07, pages 486–495, Berlin, Heidelberg, 2007. Springer-Verlag. (Cited on pages 5 and 117.)

[Trail 2005] Galen T Trail, Dean F. Anderson and Janet S. Fink. *Consumer satisfaction and identity theory: A model of sport spectator conative loyalty.* Sport Marketing Quarterly, vol. 14, no. 2, pages 98–111, 2005. (Cited on pages 4 and 116.)

[Turkle 2011] S Turkle. Alone together: Why we expect more from technology and less from each other. Basic Books, 2011. (Cited on page 54.)

[Uhrich 2010] Sebastian Uhrich, Martin Benkenstein*et al.* *Sport stadium atmosphere: formative and reflective indicators for operationalizing the construct.* Journal of Sport Management, vol. 24, no. 2, pages 211–237, 2010. (Cited on page 116.)

[Vogel 2004] Daniel Vogel and Ravin Balakrishnan. *Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users.* In Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, UIST '04, pages 137–146, New York, NY, USA, 2004. ACM. (Cited on page 61.)

[Wagner 2013] Julie Wagner, Mathieu Nancel, Sean G. Gustafson, Stephane Huot and Wendy E. Mackay. *Body-centric Design Space for Multi-surface Interaction.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pages 1299–1308, New York, NY, USA, 2013. ACM. (Cited on page 10.)

[Wang 2006] Shuo Wang, Xiaocao Xiong, Yan Xu, Chao Wang, Weiwei Zhang, Xiaofeng Dai and Dongmei Zhang. *Face-tracking As an Augmented Input in Video Games: Enhancing Presence, Role-playing and Control.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06, pages 1097–1106, New York, NY, USA, 2006. ACM. (Cited on page 63.)

[Wann 1996] Daniel L. Wann, Kathleen B. Tucker and Michael P. Schrader. *An exploratory examination of the factors influencing the organization, continuation, and cessattion of identification with sports teams.* Perceptual and motor skills, vol. 82, no. 3, pages 995–1001, 1996. (Cited on pages 4 and 116.)

[Weisz 2007] Justin D. Weisz, Sara Kiesler, Hui Zhang, Yuqing Ren, Robert E. Kraut and Joseph A. Konstan. *Watching Together: Integrating Text Chat with Video.* In Proceedings of the SIGCHI Conference on Human Factors in

Computing Systems, CHI '07, pages 877–886, New York, NY, USA, 2007. ACM. (Cited on pages 64 and 124.)

[Wittenburg 2006] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann and Han Sloetjes. *Elan: a professional framework for multimodality research.* In In Proceedings of Language Resources and Evaluation Conference (LREC, 2006. (Cited on page 41.)

Curriculum Vitae

since 07/2010	Technische Universität Darmstadt PhD candidate in the area of Human-Computer Interaction
2007 - 2010	RWTH-Aachen University, Aachen - Germany Field of Media Informatics - International Master Program
2003 - 2007	Kish University, Kish Island - Iran Bachelor of Science in Computer Engineering (software)
1999 - 2003	High School Paknejad, Tehran - Iran Mathematic-Physic diploma